

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:15:33 ; Search time 9.71143 Seconds
(without alignments)
2813.016 Million cell updates/sec

Title: US-10-006-011A-9
Perfect score: 1566
Sequence: I CERQPCQHGATCMPAGEYEF.....QELDLQHRQAQANTRPCPS 284

Scoring table: Gapop10.0, Gapext 0.5

Searched: 283366 seqs, 96191526 residues 283366
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: P1r1:*
2: P1r2:*
3: P1r3:*
4: P1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1566	100.0	4391	A38096	perlecan precursor
2	1370.5	87.5	3707	S18852	heparan sulfate pr
3	349	22.3	3375	T19821	hypothetical prote
4	343.5	21.3	1959	1 AGRT	agrin - rat
5	314.5	20.1	1328	T43060	agrin - electric r
6	301.5	19.3	1955	1 AGCH	agrin precursor -
7	283.5	18.1	1531	T42218	slit-1 protein hom
8	264	16.9	4307	T20721	hypothetical prote
9	249	15.9	5147	1 IJFTM	cadherin-related t
10	249	15.9	601	T22025	hypothetical prote
11	246	15.7	601	D89711	protein F40E10.4 l
12	246	15.7	1523	T13953	MEGF5 protein - ra
13	243.5	15.5	1025	T42626	secreted leucine-r
14	223.5	14.3	1751	1 MWHUMH	laminin alpha-2 ch
15	217	13.9	1438	T48216	neurexin III-alpha
16	217	13.9	1471	T48218	neurexin III-alpha
17	217	13.9	1578	T48216	neurexin III-alpha
18	214	13.7	2809	T30213	G-cadherin - sea u
19	207.5	13.3	3106	1 S53868	laminin alpha-2 ch
20	199	12.7	2321	T478549	notch3 protein - h
21	197.5	12.6	1530	T2	neurexin I-alpha
22	197.5	12.6	2610	T203668	hypothetical prote
23	196.5	12.5	1507	T402268	neurexin I-alpha p
24	194.5	12.4	1715	T2	neurexin II-alpha
25	191	12.2	3034	C40228	seven-pass transme
26	188.5	12.0	3075	T14119	laminin alpha-1 ch
27	185.5	11.8	3672	T23453	hypothetical prote
28	185.5	11.8	3704	T237316	probable laminin a
29	184	11.7	3097	T200021	DN-cadherin - frui

30	183	11.7	1469	2 B36665	slit protein 2 pre
31	183	11.7	1480	2 A36665	slit protein 1 pre
32	182	11.6	2318	2 S43306	notch 3 protein -
33	181.5	11.6	2531	2 S18188	notch protein homo
34	181.5	11.6	2531	2 A46019	notch-1 protein -
35	180	11.5	2531	2 T31070	notch homolog - se
36	177.5	11.3	1964	2 T09059	notch4 - mouse
37	176.5	11.3	4773	2 A56175	adhesive plaque pr
38	176	11.2	1647	2 T32934	hypothetical prote
39	175	11.2	1203	2 A49175	Notch 3 protein -
40	175	11.2	2471	2 A49128	cell-fate determin
41	174.5	11.1	2555	2 A40043	notch protein homo
42	173	11.0	1420	2 A56136	jagged protein pre
43	170	10.9	385	2 A54785	preadipocyte facto
44	170	10.9	385	2 S53718	homeotic protein d
45	169.5	10.8	1268	2 S52781	neurocan - mouse

ALIGNMENTS

RESULT 1

A38096
perlecan precursor - human
N;Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate proteoglycan
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text change 05-Nov-1999
C;Accession: A38096, S19256, S77946; A41059; A40306; B33625; A33625; A41736
R;Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J. Biol. Chem. 267, 8544-8557, 1992
A;Title: Primary structure of the human heparan sulfate proteoglycan from basement membr
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A;Reference number: A38096; MUID:92235084; PMID:1569102
A;Accession: A38096
A;Molecule type: mRNA
A;Residues: 1-4391 <MUI>
A;Cross-references: GB:M8289; NID:9184426; PIDN:AAA52700.1; PID:g184427
R;Kallunki, P.; Tryggvason, K.
J. Cell Biol. 116, 559-571, 1992
A;Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD prot
ell adhesion molecules, and epidermal growth factor.
A;Reference number: A41736; MUID:92112994; PMID:1730768
A;Accession: S19256
A;Molecule type: mRNA
A;Residues: 1-57 'D', 59-434, 'A', 436, 'E', 438-449, 'O', 451-502, 'A', 503-792, 'K', 794-908, 'R',
71-2979, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 3633-36
A;Cross-references: EMBL:X62515; NID:929469; PIDN:CAA4373.1; PID:g29470
R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
submitted to the EMBL Data Library, October 1991
A;Reference number: S77946
A;Accession: S77946
A;Molecule type: mRNA
A;Residues: 1-57 'D', 59-434, 'A', 436, 'E', 438-449, 'O', 451-502, 'A', 503-792, 'K', 794-908, 'R',
71-2979, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 3633-4
A;Cross-references: EMBL:X62515; NID:929469; PIDN:CAA4373.1; PID:g29470
R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
Genomics 11, 389-396, 1991
A;Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the 9c
A;Reference number: A41059; MUID:92120660; PMID:1685141
A;Accession: A41059
A;Molecule type: mRNA
A;Residues: 'R', 892-908, 'R', 910-1101, 'L', 1103-1132, 'L', 1134-1221, 'L', 1223-1397 <KA2>
A;Cross-references: GB:S78436; NID:9243370; PIDN:AA21121.1; PID:g243371
R;Dodge, G.R.; Kovalszky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo, F
Genomics 10, 673-680, 1991
A;Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellula
A;Reference number: A40306; MUID:91365376; PMID:1679749
A;Accession: A40306
A;Molecule type: mRNA
A;Residues: 1018-1405, 'G', 1407-1409, 'G', 1411-1465 <DOD>
A;Cross-references: GB:M64283; NID:9184424; PIDN:AAA52699.1; PID:g184425
R;Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den
J. Cell Biol. 109, 3199-3211, 1989

A;Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclona
 A;Reference number: A33625; MUID:90078352; PMID:2667294
 A;Accession: A33625
 A;Molecule type: protein
 A;Residues: 1379-1384, 'X', 1386-1388, 'X', 1390-1398 <HE2>
 A;Accession: A33625
 A;Molecule type: protein
 A;Residues: 2166-2171, 'X', 2173-2175, 'X', 2177-2185 <HE3>
 A;Note: peptide potentially matches four different regions of sequence shown
 C;Genetics:
 A;Gene: GDB:HSPG2
 A;Cross-references: GDB:126372; OMIM:142461
 A;Map position: lp36.1-lp36.1
 C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
 C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembra
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-4391/Product: perlecan #status predicted <MAT>
 F;22-193/Domain: I <DOM1>
 F;194-530/Domain: II <DOM2>
 F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F;531-1676/Domain: III <DOM3>
 F;1159-1206/Domain: laminin-type EGF-like homology <LEG>
 F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
 F;1613-1668/Domain: laminin-type EGF-like homology <LEGS>
 F;1677-3686/Domain: IV <DOM4>
 F;2007-2034/Domain: transmembrane #status predicted <TRM>
 F;3687-4391/Domain: V <DOM5>
 F;3845-3880/Domain: EGF homology <EGF1>
 F;3988-3921/Domain: EGF homology <EGF>
 F;3953-4106/Domain: laminin G repeat homology <LG2>
 F;4147-4175/Domain: EGF homology <EGF2>
 F;4149-4151/Region: motor neuron attachment (L-R-E) motif
 F;4299-4301/Region: motor neuron attachment (L-R-E) motif
 F;571,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
 F;89,554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (cov
 F;2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 1566; DB 2; Length 4391;
 Best Local Similarity 100.0%; Pred. No. 3,2e+105;
 Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGLDCEHENPCOLREPCPLHGCTCGTRCLCL 60
 DB 4108 CERQPCQHGATCMPAGEYEFQCLCRDGFKGLDCEHENPCOLREPCPLHGCTCGTRCLCL 4167
 OY 61 PFGSGPCQCGSGHGIAPSDWHLESGSGNDAPQGYGAYFHDGDFLAFPHVFSRSLPEVP 120
 DB 4168 PFGSGPCQCGSGHGIAPSDWHLESGSGNDAPQGYGAYFHDGDFLAFPHVFSRSLPEVP 4227
 OY 121 ETIELEVRTSTASGLLWQGVGEAGQKDFISLQDGHVFRYQLGSGEARLVSEDP 180
 DB 4228 ETIELEVRTSTASGLLWQGVGEAGQKDFISLQDGHVFRYQLGSGEARLVSEDP 4287
 OY 181 INDGEHRVTALREGRGSIQVDGEEELVSGRSPGNVAVNAKGSVYIGGAPDVATLTGGR 240
 DB 4288 INDGEHRVTALREGRGSIQVDGEEELVSGRSPGNVAVNAKGSVYIGGAPDVATLTGGR 4347
 OY 241 FSSGITCVKMLVLSARPGAPPQPLDLQHRACAGANTRPCPS 284
 DB 4348 FSSGITCVKMLVLSARPGAPPQPLDLQHRACAGANTRPCPS 4391

RESULT 2
 S18252
 heparan sulfate proteoglycan - mouse
 N;Alternate names: perlecan
 C;Species: Mus musculus (house mouse)
 C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
 C;Accession: S18252; A31917; S66460

R;Noonan, D.M.; Full, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaki, M.; Yamada, Y.; H
 J. Biol. Chem. 266, 22939-22947, 1991
 A;Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteogly
 adhesion molecule.
 A;Reference number: S18252; MUID:92078153; PMID:1744087
 A;Accession: S18252
 A;Molecule type: mRNA
 A;Residues: 1-3707 <NOO>
 A;Cross-references: EMBL:M77174; NID:9200295; PIDN:AAA39911.1; PID:G200296
 R;Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogeli, G.; Sasaki, M.; Yamada, Y.; Hase
 J. Biol. Chem. 263, 16379-16387, 1988
 A;Title: Identification of cDNA clones encoding different domains of the basement membra
 A;Reference number: A92680; MUID:89034110; PMID:2372708
 A;Accession: A31917
 A;Molecule type: mRNA
 A;Residues: 940-1601 <NO2>
 A;Cross-references: GB:J04054; NID:G200252; PIDN:AAA39899.1; PID:G200253
 A;Accession: B31917
 A;Molecule type: mRNA
 A;Residues: 1870-2600 <NO3>
 A;Cross-references: GB:J04055; NID:G200300; PIDN:AAA39912.1; PID:G200301
 R;Schulze, B.; Mann, K.; Battistutta, R.; Wiedemann, H.; Timpl, R.
 Eur. J. Biochem. 231, 551-556, 1995
 A;Title: Structural properties of recombinant domain III-3 of perlecan containing a glob
 A;Reference number: S66460; MUID:95377282; PMID:7649154
 A;Accession: S66460
 A;Molecule type: protein
 A;Residues: 1272-1274, 'X', 1276, 'X', 1278-1279 <SCH>
 C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
 C;Keywords: glycoprotein
 F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F;764-811/Domain: laminin-type EGF-like homology <LEG>
 F;1159-1206/Domain: laminin-type EGF-like homology <LEG7>
 F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
 F;1613-1668/Domain: laminin-type EGF-like homology <EG7>
 F;3163-3198/Domain: laminin-type EGF-like homology <LEGS>
 F;3270-3423/Domain: EGF homology <EGF>
 F;3464-3492/Domain: EGF homology <EGF7>
 F;1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 87.5%; Score 1370.5; DB 2; Length 3707;
 Best Local Similarity 86.3%; Pred. No. 4e-91;
 Matches 245; Conservative 20; Mismatches 18; Indels 1; Gaps 1;
 OY 1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGLDCEHENPCOLREPCPLHGCTCGTRCLCL 60
 DB 3425 CERQPCQHGATCMPAGEYEFQCLCRDGFKGLDCEHENPCOLREPCPLHGCTCGTRCLCL 3484
 OY 61 PFGSGPCQCGSGHGIAPSDWHLESGSGNDAPQGYGAYFHDGDFLAFPHVFSRSLPEVP 120
 DB 3485 PFGSGPCQCGSGHGIAPSDWHLESGSGNDAPQGYGAYFHDGDFLAFPHVFSRSLPEVP 3544
 OY 121 ETIELEVRTSTASGLLWQGVGEAGQKDFISLQDGHVFRYQLGSGEARLVSEDP 180
 DB 3545 ETIELEVRTSTASGLLWQGV-VREASRDKDFISLQDGHVFRYQLGSGEARLVSEDP 3603
 OY 181 INDGEHRVTALREGRGSIQVDGEEELVSGRSPGNVAVNAKGSVYIGGAPDVATLTGGR 240
 DB 3604 INDGEHRVTALREGRGSIQVDGEEELVSGRSPGNVAVNAKDIYIIGGAPDVATLTGGR 3663
 OY 241 FSSGITCVKMLVLSARPGAPPQPLDLQHRACAGANTRPCPS 284
 DB 3664 FSSGITCVKMLVLSARPGAPPQPLDLQHRACAGANTRPCPS 3707

RESULT 3
 T19821
 hypothetical protein ZC101.2e - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 23-Sep-2002

C;Accession: T19821; T19819; T19820; T27490; T27488; T27489; T27487; A47648; B47648; C47
 R;Baynes, C.
 submitted to the EMBL Data Library, March 1997
 A;Reference number: Z19182
 A;Accession: T19821
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-3375 <N1>
 A;Cross-references: EMBL:Z93375; PIDN:CAB07569.1; GSPDB:GN00020; CESP:ZC101.2e
 A;Experimental source: clone C38C
 A;Accession: T19819
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-2441, 'R', 'KRKH', '3369', 'GN', '3372-3373', 'G', '3375', 'RLRHRRNAQNGPL
 A;Cross-references: EMBL:Z93375; PIDN:CAB07567.1; GSPDB:GN00020; CESP:ZC101.2a
 A;Experimental source: clone C38C
 A;Accession: T19820
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-1694, 'H', '1883-2441', 'R', 'KRKH', '3369', 'GN', '3372-3373', 'G', '3375', 'RLRHRRNAQNGPL
 A;Cross-references: EMBL:Z93375; PIDN:CAB07568.1; GSPDB:GN00020; CESP:ZC101.2c
 A;Experimental source: clone C38C
 R;Percy, C.
 submitted to the EMBL Data Library, March 1997
 A;Reference number: Z20375
 A;Accession: T27490
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-3375 <N1>
 A;Cross-references: EMBL:Z93395; PIDN:CAB07708.1; GSPDB:GN00020; CESP:ZC101.2e
 A;Experimental source: clone ZC101
 A;Accession: T27488
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-2441, 'R', 'KRKH', '3369', 'GN', '3372-3373', 'G', '3375', 'RLRHRRNAQNGPL
 A;Cross-references: EMBL:Z93395; PIDN:CAB07706.1; GSPDB:GN00020; CESP:ZC101.2a
 A;Experimental source: clone ZC101
 A;Accession: T27489
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-1128, 1290, 'DFARNSPS', '1299', 'NSS', '1303-1304', 'R', 'RHR', '1544-1545', 'RIRVRS', '155
 A;Cross-references: EMBL:Z93395; PIDN:CAB07704.1; GSPDB:GN00020; CESP:ZC101.2b
 A;Experimental source: clone ZC101
 R;Rogalski, T.M.; Williams, B.D.; Mullen, G.P.; Moerman, D.G.
 Genes Dev. 7, 1471-1484, 1993
 A;Title: Products of the unc-52 gene in Caenorhabditis elegans are homologous to the co
 A;Reference number: A47648; MUID:93339574; PMID:8393416
 A;Accession: A47648
 A;Status: preliminary; nucleic acid sequence not shown
 A;Molecule type: DNA
 A;Residues: 1-546, 'P', '548-2198', 'D', '2290', 'NAR', '2294', 'L', '2296', 'WHATE', '2302-2303', 'V', '2305',
 A;Cross-references: GB:L13458
 A;Accession: B47648
 A;Status: preliminary; nucleic acid sequence not shown
 A;Molecule type: DNA
 A;Residues: 1-546, 'P', '548-2198', 'D', '2290', 'NAR', '2294', 'L', '2296', 'WHATE', '2302-2303', 'V', '2305',
 A;Cross-references: GB:L13458
 A;Accession: C47648
 A;Status: preliminary; nucleic acid sequence not shown
 A;Molecule type: DNA
 A;Residues: 1-546, 'P', '548-2198', 'D', '2290', 'NAR', '2294', 'L', '2296', 'WHATE', '2302-2303', 'V', '2305',
 A;Cross-references: GB:L13458
 C;Genetics:
 A;Gene: CESP:ZC101.2e; CESP:ZC101.2a; CESP:ZC101.2c; CESP:ZC101.2b
 A;Map position: 2
 A;Introns: 32/1; 134/1; 225/1; 335/2; 450/3; 739/3; 830/3; 860/2; 1064/2; 1129/1; 1158/3

2; 2613/1; 2684/1; 2757/1; 2813/3; 2863/1; 2900/3; 3084/1; 3176/1; 3250/2
 C;Superfamily: LR11 protein; laminin-type EGF-like homology; LDL receptor ligand-binding
 C;Keywords: extracellular matrix
 F;149-183/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F;190-224/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F;233-268/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F;955-1002/Domain: laminin-type EGF-like homology <LEG1>
 F;1101-1058/Domain: laminin-type EGF-like homology <LEG2>
 Query Match Similarity 22.3%; Score 349; DB 2; Length 3375;
 Best Local Similarity 31.2%; Pred. No. 4.5e-17;
 Matches 81; Conservative 41; Mismatches 102; Indels 36; Gaps 6;
 QY 30 GDLCHEEE--NPCQLRPFCLHGGTTCOCTR-----CLCLPGFSPRCQGGHGAESDWH 82
 DB 3129 GDVYSTQEPNNIC-ANSTCGMNGQVPRNMTHTYCEKLYYDGTCSLFF----- 3177
 QY 83 LEGSGNDAPQCYGAYFHDDGFLAAPPGHVRSRSLPVEPTELEEVRTSTASGLLWQGV 142
 DB 3178 -----PIEHAARFDGDAFIELSDEFPHLTSEKDEIVAFKFKTEQQNGVLLWQG-Q 3227
 QY 143 VGEAGQGKDFISLGLQDHLVFRYQLSGEARLVSEDDPINDGEWHRVYALFREGRRGSIQV 202
 DB 3228 RPTVQOMEDIYSVGVNGLHFSYELGGAAHLISERVDGKESVRFERKRGEGOMRI 3287
 QY 203 DGEELVGRSPGNVAVNAKGSVVIIGAPDVAITLGTGRFSSGITCCVKNLVLHARPGAP 262
 DB 3288 DNYREVDGRSTGAILAMLNVLDGNIIFVGGVDDISKATGGLFSLNFFVGGCIADVELNGVK----- 3343
 QY 263 PPOPLDLQHRQAQANTRPC 282
 DB 3344 ----LDLMTAIDGNKVKPC 3359
 RESULT 4
 agrin - rat
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
 C;Accession: JH0399; A38856
 R;Rupp, F.; Payan, D.G.; Magill-Soic, C.; Cowan, D.M.; Scheller, R.H.
 Neuron 6, 811-823, 1991
 A;Title: Structure and expression of a rat agrin.
 A;Reference number: JH0399; MUID:91222570; PMID:1851019
 A;Accession: JH0399
 A;Molecule type: mRNA
 A;Residues: 1-1779;1799-1959 <RUP>
 A;Cross-references: GB:M4780; NID:g202798; PIDN:AAA40703.1; PID:g202800
 A;Experimental source: embryonic spinal cord
 A;Note: It is uncertain whether Met-1, Met-18, or Met-24 is the initiator
 R;Rupp, F.; Oezcelik, T.; Linial, M.; Peterson, K.; Francke, U.; Scheller, R.
 J Neurosci 12, 3535-3544, 1992
 A;Title: Structure and chromosomal localization of the mammalian agrin gene.
 A;Reference number: A38856; MUID:92407628; PMID:1326608
 A;Accession: A38856
 A;Molecule type: mRNA
 A;Residues: 1780-1798 <RU2>
 A;Cross-references: GB:S44194
 C;Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine r
 C;Comment: 90% of rat embryonic transcripts encode the variant labeled below as form 3. I
 ycholine receptor clustering activity.
 C;Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repeat
 C;Keywords: alternative splicing; duplication; Glycoprotein; neuromuscular junction
 F;1-1787/1799-1959/Product: agrin, form 1 #status predicted <AG1>
 F;1-1779/1799-1959/Product: agrin, form 4 #status predicted <AG4>
 F;1-1779/1788-1959/Product: agrin, form 3 #status predicted <AG3>
 F;1-1143,1153-1959/Product: agrin, form 5 #status predicted <AG5>
 F;22-50/Region: hydrophobic
 F;88-137/Domain: Kazal proteinase inhibitor homology <KPI1>
 F;163-212/Domain: Kazal proteinase inhibitor homology <KPI2>
 F;236-284/Domain: Kazal proteinase inhibitor homology <KPI3>
 F;307-356/Domain: Kazal proteinase inhibitor homology <KPI4>

F:381-429/Domain: Kazal proteinase inhibitor homology <KPI5>
F:446-494/Domain: Kazal proteinase inhibitor homology <KPI6>
F:511-559/Domain: Kazal proteinase inhibitor homology <KPI7>
F:540-542/Region: motor neuron attachment (L-R-E) motif
F:596-645/Domain: Kazal proteinase inhibitor homology <KPI8>
F:688-739/Domain: laminin-type EGF-like homology <LE1>
F:742-786/Domain: laminin-type EGF-like homology <LE2>
F:814-864/Domain: laminin-type EGF-like homology <LE3>
F:869-992/Region: serine/threonine-rich
F:1094-1086/Region: motor neuron attachment (L-R-E) motif
F:1147-1215/Domain: serine/threonine-rich
F:1287-1442/Domain: laminin G repeat homology <LG1>
F:1444-1476/Domain: EGF homology <EG2>
F:1483-1515/Domain: EGF homology <EG3>
F:1555-1706/Domain: laminin G repeat homology <LG2>
F:1713-1747/Domain: EGF homology <EG4>
F:1807-1959/Domain: laminin G repeat homology <LG3>
F:97-116, 105-137, 171-191, 180-212, 244-263, 252-284, 316-335, 324-356, 389-408, 397-429, 454-473
-1476, 1483-1494, 1488-1504, 1506-1515/Disulfide bonds: #status predicted
F:145,672,827,957/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 21.3%; Score 333.5; DB 1; Length 1959;
Best Local Similarity 36.0%; Pred. No. 3.4e-16;
Matches 94; Conservative 28; Mismatches 108; Indels 31; Gaps 10;

QY 1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGLDCEHENPCOLRPPCLHGGTCQ-----GT 55
Db 1444 CUPNFCGALCQALEAGMFLCQCPGFRGPTCADEKSPCO-PNFCGAAACRVLSSGGA 1502

QY 56 RCLCLPFGSGPRCQQSGHGIAESDWHLEGGGNDAPGQYGFHDDGFLAAPP-GHVFPSR 114
Db 1503 KCECPVETIEVRSSTAGSLLMQGVGEAGQKDFISLGLQDGHVFRYQLGSGEAR 174

QY 115 SLPEVPEVTEIEVRSSTAGSLLMQGVGEAGQKDFISLGLQDGHVFRYQLGSGEAR 174
Db 1548 DLGE-KMALEWVFLARGSGLLYNGQKTD--GKG-DFVSLAHRHLEFCYDLKGNAAV 1603

QY 175 LVSEDPINDGEHWRVTLRREGRRGSIQVDGEEVLVSGRSPG---PNVAVNAKGSVYIGGA 230
Db 1604 IRSKEPIALGTVWRVFLERNGRKGALQVGDGPRVLGSPKSRKVPHTMLNKLPLYIGGA 1663

QY 231 PDVAIILT-GGRFSSGITGVCK 250
Db 1664 PDFSKLARGAAYSSGFSGVIQ 1684

RESULT 5
T43060
agrin - electric ray (Discopyge ommata) (fragment)
C:Species: Discopyge ommata
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 17-Nov-2000
R:Smith, M.A.; Magill-Solc, C.; Rupp, F.; Yao, Y.M.M.; Schilling, J.W.; Snow, P.; McMaha
submitted to the EMBL Data Library, September 1992
A:Reference number: Z22308
A:Accession: T43060
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1328 <SMI>
A:Cross-references: EMBL:L01423; NID:g213102; PID:g213103; PIDN:AAA49224.1
C:Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C:Keywords: glycoprotein; neuromuscular junction

Query Match 20.1%; Score 314.5; DB 2; Length 1328;
Best Local Similarity 31.4%; Pred. No. 5.4e-15;
Matches 85; Conservative 44; Mismatches 105; Indels 37; Gaps 12;

QY 1 CERQPCQHGATCMPAGEYEFQCLCRDGFKGLDCEHENPCOLRPPCLHGGTCQ-----GT 55
Db 830 CSFNPCVNRKCKHKEAEMFCHESVGEFSGPTCADKHNPDC-PNFCQAGNCWVLPGGGS 888

QY 56 RCLCLPFGSGPRCQQSGHGIAESDWHLEGGGNDAPGQYGFHDDGFLAAPP-GHVFPSR 114

Db 889 KCCEPMGREGELCER-----VSEAE-----QQQKFAIPEFNGLSYLEMNGLIHTF-- 933
QY 115 SLPEVPEVTEIEVRSSTAGSLLMQGVGEAGQKDFISLGLQDGHVFRYQLGSGEAR 172
Db 934 -VSDLLQKLSMEVIFLAKDPNGMIFYNGKTD--GRG-DFVSLAHRDGLYLFKFKYDLGKGA 989

QY 173 ARLVSEDPINDGEHWRVTLRREGRRGSIQVDGEEVLVSGRSP---GNVAVNAKGSVYIG 228
Db 990 AVLRKAPILPNNVNVTVERNRKGKMLKINKDELVSSEPKSRKAPHTALNKLKEAFVVG 1049

QY 229 GAPD---VATLTCGRFSSGITGVCKVNLVHLS 256
Db 1050 GAPDFNKPARAAG--IISGFTGAIQKLSLKS 1078

RESULT 6
AGCH
agrin precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
C:Accession: JH0591; A38857; E38857; I50692
R:Tsimm, K.W.K.; Ruegg, M.A.; Escher, G.; Kroeger, S.; McMahan, U.J.
Neuron 8, 677-689, 1992
A:Title: cDNA that encodes active agrin.
A:Reference number: JH0591; MUID:92232297; PMID:1314620
A:Accession: JH0591
A:Molecule type: mRNA
A:Residues: 1-1955 <TSI>
A:Cross-references: GB:M94271; NID:g211120; PIDN:AAA48585.1; PID:g211121
A:Experimental source: brain
R:Tsimm, M.A.; Tsimm, K.W.K.; Horton, S.E.; Kroeger, S.; Escher, G.; Gensch, E.M.; McMaha
Neuron 8, 691-699, 1992
A:Title: The agrin gene codes for a family of basal lamina proteins that differ in funct
A:Reference number: A38857; MUID:92232298; PMID:1314621
A:Contents: alternative splicing
A:Accession: A38857
A:Molecule type: mRNA
A:Residues: 1132-1793; 1795-1955 <RU2>
A:Cross-references: GB:M97371
A:Accession: B38857
A:Molecule type: mRNA
A:Residues: 1221-1647; 1652-1783; 1794-1955 <RU3>
A:Cross-references: GB:M97372
A:Note: translation of the nucleotide sequence is not complete
R:Thomas, W.S.; O'Dowd, D.K.; Smith, M.A.
Dev. Biol. 158, 523-535, 1993
A:Title: Developmental expression and alternative splicing of chick agrin RNA.
A:Reference number: I50692; MUID:93345745; PMID:8393816
A:Accession: I50692
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: SHLSNEIPA, 1784-1795 <THO>
A:Cross-references: EMBL:U07271; NID:9459665; PIDN:AAA16788.1; PID:g459666
C:Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine r
C:Comment: Alternative splicing produces two inactive proteins: agrin-related protein 1
C:Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
C:Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
F:1-38/Domain: signal sequence #status predicted <SIG>
F:139-1955/Product: agrin #status predicted <MA>
F:139-1793, 1795-1955/Product: agrin-related protein 1 #status predicted <AG1>
F:139-1647, 1652-1793, 1794-1955/Product: agrin-related protein 2 #status predicted <AG2>
F:77-126/Domain: Kazal proteinase inhibitor homology <KPI1>
F:152-201/Domain: Kazal proteinase inhibitor homology <KPI2>
F:225-273/Domain: Kazal proteinase inhibitor homology <KPI3>
F:295-344/Domain: Kazal proteinase inhibitor homology <KPI4>
F:370-418/Domain: Kazal proteinase inhibitor homology <KPI5>
F:433-483/Domain: Kazal proteinase inhibitor homology <KPI6>
F:500-548/Domain: Kazal proteinase inhibitor homology <KPI7>
F:584-633/Domain: Kazal proteinase inhibitor homology <KPI8>
F:675-726/Domain: laminin-type EGF-like homology <LE1>
F:729-773/Domain: laminin-type EGF-like homology <LE2>
F:801-851/Domain: Kazal proteinase inhibitor homology <KPI9>

F:856-995/Region: serine/threonine-rich
 F:1150-1219/Region: serine/threonine-rich
 E:1233-1264/Domain: EGF homology <EG1>
 E:1294-1448/Domain: laminin G repeat homology <LG1>
 F:1429-1433/Region: motor neuron attachment (L-R-E) motif
 F:1450-1482/Domain: EGF homology <EG2>
 F:1489-1521/Domain: EGF homology <EG3>
 F:1560-1711/Domain: laminin G repeat homology <LG2>
 F:1718-1751/Domain: EGF homology <EG4>
 F:1803-1955/Domain: laminin G repeat homology <LG3>
 E:86-105_94-126_150-180_169-201_233-252_241-273_304-323_312-344_378-397_386-418_443-462_482_1489-1500_1494-1510_1512-1521/Disulfide bonds: #status predicted
 F:390,659,764,814/Binding site: carbohydrate (Asn) #status predicted

Query Match 19.3%; Score 301.5; DB 1; Length 1955;
 Best Local Similarity 27.4%; Pred. No. 7.1e-14;
 Matches 87; Conservative 38; Mismatches 108; Indels 85; Gaps 13;

Qy 1 CERQPCQHGATCMPAGEVEYFQCLCRDGFKGLDCHHENPCQLRPECLHGTC-----QGT 55
 Db 1450 CHNFCRHGASCHVKEAMFHCECLHSYTGPTCADERNPCD-PTPCHISATCLVLEGG 1508
 Qy 56 RCLCLPFGSPRCQSGHGIAESDWHLEGGNDAPFGQYGAFFHDDGFLAFLPFGHVFSS 115
 Db 1509 MCACPMRGEGFCER-----VTEQD-----HT 1530
 Qy 116 LPEVPE-----TIELEVRTASGLLLWQGEVGEAGQKDFI 153
 Db 1531 MPFLPEFNFGSYLELNGQLTLFLTRQMSMEVFLAKSPGMIFYGQKTD--GKG-DFV 1587
 Qy 154 SLGLQDGHVFRYQLGGEARLVSEDPNGEHRVTRALRGRGSIQVDGEBELVGRSP 213
 Db 1588 SLALHDGYLERYDLGRGAVALRSKEFVPLNTWISVLLRSRGRKGVNRINNGERVMGESP 1647
 Qy 214 G----PNVAIVNAKSVYIGGAPDVATLT-GRFSSGITGCVKLVLSRAPPGPPLD 268
 Db 1648 KSRKVPFAFLNKEFFVYGGAPDFSKLARAASITSPYGAQRISI----KGVP----LLK 1700
 Qy 269 LQH-RAQAGANT--RPC 282
 Db 1701 EQHLRSVEIETFRAHFC 1718

RESULT 7
 T42218
 slit-1 protein homolog - rat
 N:Alternate names: MEGF4 protein
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Aug-2002
 C:Accession: T42218
 R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
 Genomics 51, 27-34, 1998
 A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
 A:Reference number: Z14126; MUID:98360089; PMID:9693030
 A:Accession: T42218
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1531 <NAK>
 A:Cross-references: EMBL:AB011530; NID:G3449289; PIDN:BA32460.1; PID:G3449290
 A:Experimental source: strain Sprague-Dawley; brain
 C:Genetics:
 A:Gene: MEGF4
 C:Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein 2

Query Match 18.1%; Score 283.5; DB 2; Length 1531;
 Best Local Similarity 27.9%; Pred. No. 1.1e-12;
 Matches 81; Conservative 48; Mismatches 92; Indels 69; Gaps 14;

Qy 1 CERQPCQHGATCMPAGEVEYFQCLCRDGFKGLDCHHENPCQLRPECLHGTC--QGT 55
 Db 1087 KDHQCCQNGAQCQVDEIN-SYACLRAGSYGQJCEIPPAFRNSCEGTE-CQNGANCVDGSS 1144
 Qy 56 R--CLCLPFGSPRCQSGHGIAESDWHLEGGNDAPFGQYGAFFHDDGFLAFLPFGHVFSS 113

Db 1145 RPVQCCLFPGFGECEKLLSNVFNDRDYLQ-----FTD-----1178
 Qy 114 RSLPEVPE-TIELEVRTASGLLLWQGEVGEAGQKDFISLGLQDGHVFRYOLGS-G 171
 Db 1179 --LQNWPRANITLQVSTAEDNGILLXG-----DNDHIAVELYQGHVRSVYDPGSSYP 1228
 Qy 172 EARLVSEDPNGEHRVTRALRGRGSIQVDGEBELVGRSGPNVAIVNAKGSVYIGGAP 231
 Db 1229 SSAIYSAEITNDQPHITVELVTFDQMVNLSIDGSGPMTDMDFGKRYITLNSRAPLYVGGMP 1288
 Qy 232 -DVAT-----LTGGRFSSGITGCVKLVLS-----APPGAPP 263
 Db 1289 VDVNSAARLWQILNGTFSH---GCIRNLVINNELODFTKTQMKPGVVP 1334

RESULT 8
 T20721
 hypothetical protein F25F2.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 19-May-2000
 C:Accession: T20721; T21343; T23842
 R:Ainscough, R.
 submitted to the EMBL Data Library, August 1994
 A:Reference number: Z19314
 A:Accession: T20721
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-4307 <WIL>
 A:Cross-references: EMBL:Z35662; PIDN:CAA84721.1; GSPDB:GN00021; CESP:F25F2.2
 R:Ainscough, R.
 submitted to the EMBL Data Library, August 1994
 A:Reference number: Z19410
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-4307 <W12>
 A:Cross-references: EMBL:Z35599; PIDN:CAA84661.1; GSPDB:GN00021; CESP:F25F2.2
 R:Sulston, J.
 submitted to the EMBL Data Library, June 1994
 A:Reference number: Z19806
 A:Accession: T23842
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-4307 <W13>
 A:Cross-references: EMBL:Z34802; PIDN:CAA84339.1; GSPDB:GN00021; CESP:F25F2.2
 A:Experimental source: clone M88
 C:Genetics:
 A:Gene: CESP:F25F2.2
 A:Map position: 3
 Query Match 16.9%; Score 264; DB 2; Length 4307;
 Best Local Similarity 26.4%; Pred. No. 8.6e-11;
 Matches 81; Conservative 45; Mismatches 99; Indels 82; Gaps 13;

Qy 21 QCLCRDGFKGLDCHHENPC-----QLRPECLHGTCQCLRCLCLPFGSPRCQ--- 69
 Db 3670 QCLCPDGFGRKCEVETNCAKSPCQWQLCIPSVHNSTYE---CVCPLGMEGDKCVSFS 3726
 Qy 70 -QSGHGIAESDWHLEGGNDAPFGQYGAFFHDDGFLAFLPFGHVFSSRSLPEVPE---IEL 125
 Db 3727 CQNDGKCLREBELSVGGD-----YFELSLSNELETRMELSI 3763
 Qy 126 EVRTSTASGLLLWQGEVGEAGQK-DFISLGLQDGHVFRYOLGSGBEARLVSEDPNG 184
 Db 3764 ELKTTTNGIIMW-----SRGKRDFHMLRLVNGTPEYHWNAGTGTGIVTSKTSVVDG 3815
 Qy 185 EHRVTRALRGRGSIQVDGEBELVGRSGPNVAIVNA-----KGSVYIGGAPDVATLTGGRF 241

Db 3816 QWHRIASRRQRTRMTVDDEDLQEAFFSFIGSTVINLHRYSKLVLG-----AKVDDGEL 3870
 QY 242 SSGITGCVK-----NLVLSARPG-----APPQPLDLQHR-----AQAGAN 278
 Db 3871 TDGVSACRFTTIVDGMKVLKTRQGMKLFQAGPFGSALTSFPCNDLPCQHAGTCISQKSH 3930
 QY 279 TR-PCPS 284
 Db 3931 FKCECPs 3937

RESULT 9
 IUFFTM
 cadherin-related tumor suppressor precursor - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Feb-1997
 C:Accession: A41087; B41087
 R:Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Biessmann, H.; Bryant, P.J.; Goodman, C.S.
 Call 57, 853-868, 1991
 A>Title: The fat tumor suppressor gene in Drosophila encodes a novel member of the cadhe
 A:Reference number: A41087; PMID:92069752; PMID:1959133
 A:Accession: A41087
 A:Molecule type: mRNA
 A:Residues: 143-485;1279-5147 <MAH>
 A:Cross-references: GB:M80537
 A:Accession: B41087
 A:Molecule type: DNA
 A:Residues: 1-142;487-1278 <MA2>
 A:Cross-references: GB:M80537
 A>Note: 1229-Gly and 1233-Ser were also found
 C:Genetics:
 A:Gene: fat
 C:Cross-references: FlyBase:FBgn0001075
 C:Superfamily: cadherin-related tumor suppressor; cadherin repeat homology; EGF homology
 C:Keywords: calcium binding; cell adhesion; duplication; transmembrane protein
 F:1-35/Domain: signal sequence #status predicted <SIG>
 F:36-5147/Product: cadherin-related tumor suppressor #status predicted <MAT>
 F:36-4883/Domain: extracellular #status predicted <EXT>
 F:51-156/Domain: cadherin repeat homology <CR1>
 F:159-270/Domain: cadherin repeat homology <CR2>
 F:271-382/Domain: cadherin repeat homology <CR3>
 F:390-494/Domain: cadherin repeat homology <CR4>
 F:497-599/Domain: cadherin repeat homology <CR5>
 F:602-708/Domain: cadherin repeat homology <CR6>
 F:718-822/Domain: cadherin repeat homology <CR7>
 F:831-942/Domain: cadherin repeat homology <CR8>
 F:948-1049/Domain: cadherin repeat homology <CR9>
 F:1052-1153/Domain: cadherin repeat homology <CR10>
 F:1156-1278/Domain: cadherin repeat homology <CR11>
 F:1281-1384/Domain: cadherin repeat homology <CR12>
 F:1387-1489/Domain: cadherin repeat homology <CR13>
 F:1492-1601/Domain: cadherin repeat homology <CR14>
 F:1607-1713/Domain: cadherin repeat homology <CR15>
 F:1717-1823/Domain: cadherin repeat homology <CR16>
 F:1826-1932/Domain: cadherin repeat homology <CR17>
 F:1925-2027/Domain: cadherin repeat homology <CR18>
 F:2028-2167/Domain: cadherin repeat homology <CR19>
 F:2169-2278/Domain: cadherin repeat homology <CR20>
 F:2281-2384/Domain: cadherin repeat homology <CR21>
 F:2387-2491/Domain: cadherin repeat homology <CR22>
 F:2494-2596/Domain: cadherin repeat homology <CR23>
 F:2599-2703/Domain: cadherin repeat homology <CR24>
 F:2707-2810/Domain: cadherin repeat homology <CR25>
 F:2813-2913/Domain: cadherin repeat homology <CR26>
 F:2915-3013/Domain: cadherin repeat homology <CR27>
 F:3014-3124/Domain: cadherin repeat homology <CR28>
 F:3127-3229/Domain: cadherin repeat homology <CR29>
 F:3232-3334/Domain: cadherin repeat homology <CR30>
 F:3337-3439/Domain: cadherin repeat homology <CR31>
 F:3442-3545/Domain: cadherin repeat homology <CR32>
 F:3548-3651/Domain: cadherin repeat homology <CR33>
 F:3654-3756/Domain: cadherin repeat homology <CR34>

F;3954-4010/Domain: EGF homology <EG1>
 F;4017-4048/Domain: EGF homology <EG2>
 F;4056-4089/Domain: EGF homology <EG3>
 F;4096-4127/Domain: EGF homology <EG4>
 F;4584-4609/Domain: transmembrane #status predicted <TM>
 F;4610-5147/Domain: intracellular #status predicted <INT>
 Query Match 16.9%; Score 264; DB 1; Length 5147;
 Best Local Similarity 29.3%; Pred. No. 1e-10;
 Matches 93; Conservative 39; Mismatches 103; Indels 82; Gaps 19;
 QY 1 CERQPCQHGATCM--PAGEYEFCCLCRDGFKGDLCEHEENPCQLREPCLLHGTC-----QG 54
 Db 4056 CYSKPCRNCGSCORSPDSSYF-CLCRPGRNQCSEVSDSCR-FNPLHGLGLCVSLKPG 4113
 QY 55 TRCLCLPGRSGPRCQCGSHGIAESDWHLEGGSGNDAPQYGAIFYHDDGFLAAPPQHVFSR 114
 Db 4114 YKNCNTPGRYGRHCRFS-----YG--FQPLSYMFP----- 4143
 QY 115 SLPEVPTIELEVRTSTASGLLWQ-GVEVGRAGQKDFISLQDGHVLFVRYQLGSGEA 173
 Db 4144 ALDVTINDISIVFATTKPNSLLLYNGM--OSGGRSDFLAIELVHDRAYF-----SSGGA 4196
 QY 174 R-----LVSEDPINDGEWHRTALREGRGSIQV-----DGE---ELVSGRSP-----G 214
 Db 4197 RTAISTVIAGRNLAADGWHKVTATNRGRVMSLSVAKCADSGVCTECLFGSDSCLYADEVG 4256
 QY 215 PNVAVN-AKGSVYIGG---APDVATLTGGRFSGSITGCVKLVLSHARPGAPPQPLDLQ 270
 Db 4257 FVGTLFNFKQPLMIGLSSADPILERPQGVHSDLLVGC-----LHSVHIGG---RALNLS 4308
 QY 271 HRAQ-----AGANTRFC 282
 Db 4309 SPLQQRGILAGCNRQAC 4325

RESULT 10
 T22025
 Hypothetical protein F40E10.4 - Caenorhabditis elegans (fragment)
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T22025
 R:Smye, R.
 submitted to the EMBL Data Library, February 1996
 A:Reference number: Z19503
 A:Accession: T22025
 A:Status: preliminary; translated from GB/EMBL/DDBU
 A:Molecule type: DNA
 A:Residues: 1-601 <WIL>
 A:Cross-references: EMBL:Z69792; PIDN:CRA93668.1; GSPDB:GN00028; CESP:F40E10.4
 A:Experimental source: clone F40E10
 C:Genetics:
 A:Gene: CESP:F40E10.4
 A:Map position: X

Query Match 15.9%; Score 249; DB 2; Length 601;
 Best Local Similarity 26.7%; Pred. No. 1.3e-10;
 Matches 84; Conservative 48; Mismatches 115; Indels 68; Gaps 17;
 QY 5 PCQHGATMPA-GEYEFQCLCRDGFKDLCEHEENPCQLREPCLLHGTC---QGTCLC 59
 Db 186 PCENNGKCIPIGSIY--SCMSPGFTGNCTWIDCKNVE-CQNGSCVDGILLSYDCLC 242
 QY 60 LPFGSG-----PRCQ---GSGHGIA---ESDWHL---EGSGNDAPGQ 94
 Db 243 RPYAGQYCEIPFMMWMEYKTDACQCSAGCQGCQCVASONSDFTKCHEFGSGPSCDRQ 302
 QY 95 YGAYFHDDG-FLAPGVHFSRSLPEVPE-TIELEVRTSTASGLLWQGVGEAGQKDF 152
 Db 303 MSVGFKNPGAYLALD-----PLASDGTITWLRITSKIGILLIYVGD-----HF 346
 QY 153 ISLGLQDGHVLFVRYQLGSGEA-RLVSEDPINDGEWHRTALREGRGSIQVDFE--ELVS 209

Db 347 VSAELYDGRVXLYVYIGNFPASHMVSSVKVNDGLPHRISIRTSEKCFQLQDKNPQVIVE 406
 Qy 210 GRSFGPNVAVNAKGSVYIGGAP-----DVATLTGGRFSSGITCVKLVLSRPGAPPP 264
 Db 407 NSGKSDQITKEMLYIGGLPIEKSDAKRFRHVKNSESLKGCISITINEV-----459
 Qy 265 QPLDLQHPAQAGANT 279
 Db 460 -PINLQ-QALENVNT 472

RESULT 11
 D89711
 Protein F40E10.4 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
 C:Accession: D89711
 R:Anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
 A:Reference number: A75000; MUID:99069613; PMID:1981916
 A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
 A:Accession: D89711
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-601 <STO>
 A:Cross-references: GB:chr_X; PIDN:CAA93668.1; PID:G3877014; GSPDB:GN00028; CESP:F40E10.
 C:Genetics:
 A:Gene: F40E10.4
 A:Map position: X

Query Match 15.9%; Score 249; DB 2; Length 601;
 Best Local Similarity 26.7%; Pred. No. 1.3e-10;
 Matches 84; Conservative 48; Mismatches 115; Indels 68; Gaps 17;

Qy 5 PCQHGATMPA-GEYEFQCLCRDGFKDLCEHNPQLREPLHGGTC-----QGTRCLC 59
 Db 186 PCENNGKCIPIKNGSY--SCMCGPFTGNCTETNDDCKNVE--CONGGSCVDGSLSYDCLC 242
 Qy 60 LPFGSG-----PRCQ-----GSGHGIA---ESDWHL---EGSGNDAPQ 94
 Db 243 RPYAGVQCEYPPMMEYQKTDACQCSAGCGEYASQNSSDFTCKCHEGFGSPCDRQ 302
 Qy 95 YGAVPHDDG-FLAPGHVFSRSLPEVPE-TIELEVRTSTASGLLWQGVVEBAGQCKDF 152
 Db 303 MSVGFKNFGAYLAD-----FLASDGTITMTLRTTSKIGILLYYGD-----HF 346

Qy 153 ISLGLDGHVFRYQLGSGEA-RLVSDPINDGSEHVRVTLREGRGSIQVDE--ELVS 209
 Db 347 VSAELYDGRVXLYVYIGNFPASHMVSSVKVNDGLPHRISIRTSEKCFQLQDKNPQVIVE 406

Qy 210 GRSFGPNVAVNAKGSVYIGGAP-----DVATLTGGRFSSGITCVKLVLSRPGAPPP 264
 Db 407 NSGKSDQITKEMLYIGGLPIEKSDAKRFRHVKNSESLKGCISITINEV-----459

Qy 265 QPLDLQHPAQAGANT 279
 Db 460 -PINLQ-QALENVNT 472

RESULT 12
 T13953
 MGF5 protein - rat
 N:Alternate names: slit protein homolog
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 16-Aug-2002
 C:Accession: T13953
 R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
 Genomics 51, 27-34, 1998
 A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
 A:Reference number: Z14126; MUID:96360089; PMID:96593030
 A:Accession: T13953

A:Status: preliminary; translated from GB/EMBL/DBS
 A:Molecule type: mRNA
 A:Residues: 1-1523 <NAK>
 A:Cross-references: EMBL:AB011531; NID:G3449291; PIDN:BA032461.1; PID:G3449292
 C:Genetics:
 A:Gene: MGF5
 C:Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein r

Query Match 15.7%; Score 246; DB 2; Length 1523;
 Best Local Similarity 27.1%; Pred. No. 5.8e-10;
 Matches 79; Conservative 40; Mismatches 106; Indels 66; Gaps 13;

Qy 1 CERQPCQHGATMPAGEYEFQCLCRDGFKDLCEH-----EENPCQLREPLHGGTC- 52
 Db 1078 CVAHKCRHGAQCDAVN-GYTCICPQGFSLFCEHPPMVLQTSFCDQYE-CQNGAQCI 1135
 Qy 53 --QGTRCLCLPGRSGRCCQGGSGHGAESDMELEGSGGNDAPGQYGAFFHDDGFLA 109
 Db 1136 VVQEPFCRCPFPAGRCCKLITVNFVKDYSVELASAKVRP----- 1178
 Qy 110 HVFSRSLPEVPEITIELEVRTSTASGLLWQGVVEBAGQCKDFISLGLDGHVFRY-OL 168
 Db 1179 -----QANTSLQVATDKDNGILLYK-----DNDPLALELYQGHVRLYDLSL 1220
 Qy 169 GSGEARLVSDPINDGSEHVRVTLREGRGSIQVD-GEELVSR-SPGPNVAVNAKGSVY 226
 Db 1221 SSPFTYVSYVTVNDGQFVHVELMNLQTLVVDKAPKSLGKQKQPAVGIN--PLY 1278
 Qy 227 IGGAPDVATUG-----GRFSSGITCVKLVLSRPG--APPQPL 267
 Db 1279 LGGIP---TSTGLSALRQAGDRPLGFGHGTIHEVRINNELQDFKALPPOS 1326

RESULT 13
 T42626
 secreted leucine-rich repeat-containing protein Slit2 - mouse (fragment)
 N:Alternate names: neurogenic extracellular slit protein
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 16-Aug-2002
 C:Accession: T42626
 R:Holmes, G.P.; Negus, K.; Burridge, L.; Raman, S.; Algar, E.; Yamada, T.; Little, M.H.
 Mech. Dev. 79, 57-72, 1998
 A:Title: Distinct but overlapping expression patterns of two vertebrate slit homologs im
 A:Reference number: 222177; MUID:99279238; PMID:10349621
 A:Accession: T42626
 A:Status: preliminary; translated from GB/EMBL/DBS
 A:Molecule type: mRNA
 A:Residues: 1-1025 <HOL>
 A:Cross-references: EMBL:AF074960; NID:G4151258; PID:G4151259; PIDN:AAD04345.1
 C:Genetics:
 A:Gene: Slit2
 C:Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein r

Query Match 15.5%; Score 243.5; DB 2; Length 1025;
 Best Local Similarity 24.8%; Pred. No. 5.8e-10;
 Matches 92; Conservative 45; Mismatches 103; Indels 131; Gaps 18;

Qy 1 CERQPCQHGATC-MPAGEYB-FOCLCRDGFKDLCE-----HEENPC----- 40
 Db 457 CTSNPKCHGGTCHLKEGENAGFWCTCADQFEGENVEVWDDCEDDCCENNSTCVGINNY 516
 Qy 41 -----QLREPLHGGTC-----QGTRCLCLPFGSGP----- 66
 Db 517 TCLCPPEYTGELCEEKLDLFAQDLNFCQDHSKILTPFGFKCDCTPGYIGEHCDIDFDDC 576
 Qy 67 ---RCQGGSGHGAESDWHL---EGSGG-----NDA----- 91
 Db 577 QDNKCKNGAHCCTAVNGYTCVCEPGYSGLFCFSPFWLPRTPSPCDNFCQNGAQCIIRI 636
 Qy 92 -----PQYGA-----YFDDGFLAPGHVFSRSLPEVPE-TIELEVRTSTAS 133
 Db 637 NEFICQCLPYLGEKCKELVSNVFNKESYLQIP-----SAKVRPQTNIHQIATDSDS 690

Qy 134 GLLMQGVEGEAGQKDFSLGQDGHVFRYQLGSGEARLV-SBDPINDGEMHWRTAL 192
 |::|::|
 Db 691 GILLYKG-----DKDHLAVELRGRVRAAS YDTGSHPEASALYSVETINDGNFHVIVELL 742
 |::|::|
 Qy 193 REGRRGSIQVDG--EELVSGRSQPNVAVNAKGSVVIIGAP--DVATL--TGRSPSSGI 245
 |::|::|
 Db 743 TLDSSLSSLVDSGSPKVTNLS--KQSTLNFDSPILVGGSPGNVAVASLRQAPQNGTSF 800
 |::|::|
 Qy 246 TGCVKNLVLSHS 256
 |::|::|
 Db 801 HGCIRNLVINS 811
 |::|::|

RESULT 14
 MMEHUMH
 laminin alpha-2 chain - human (fragment)
 A/Alternate names: laminin M chain; merosin heavy chain
 C/Species: Homo sapiens (man)
 C/Date: 30-Sep-1991 #sequence_revision 21-Aug-1998 #text_change 10-Dec-1999
 C/Accession: X0082; A35899; A38970; S14461
 R/Hori, H.; Kanamori, T.; Mizuta, T.; Yamaguchi, N.; Liu, Y.; Nagai, Y.
 J. Biochem. 116, 1212-1219, 1994
 A/Title: Human laminin M chain: Epitope analysis of its monoclonal antibodies by immunoblotting
 A/Reference number: X0082; MUID:95221315; PMID:7535762
 A/Accession: X0082
 A/Molecule type: mRNA
 A/Residues: 1-1751 <HOR>
 A/Experimental source: placenta
 R/Ehrig, K.; Leivo, I.; Argraves, W.S.; Ruoslahti, E.; Engvall, E.
 Proc. Natl. Acad. Sci. U.S.A. 87, 3264-3268, 1990
 A/Title: Merosin, a tissue-specific basement membrane protein, is a laminin-like protein
 A/Reference number: A35899; MUID:90238994; PMID:2185464
 A/Accession: A35899
 A/Status: nucleic acid sequence not shown
 A/Molecule type: mRNA
 A/Residues: 'V', 623-1751 <EHR1>
 A/Cross-references: EMBL:M59832
 A/Accession: A38970
 A/Molecule type: protein
 A/Residues: 1368-1384:1389-1406;1593-1607 <BHR2>
 A/Note: the sequence from fig. 1 is inconsistent with that from fig. 2 in lacking 1599-1607
 R/Ehrig, K.; Leivo, I.; Argraves, S.W.; Ruoslahti, E.; Engvall, E.
 submitted to the EMBL Data Library, December 1990
 A/Description: The tissue-specific basement membrane protein merosin is a laminin-like F
 A/Reference number: S14461
 A/Accession: S14461
 A/Molecule type: mRNA
 A/Residues: 'V', 623-1264, 'R', 1266-1751 <LE1>
 A/Cross-references: EMBL:M59832; NID:G187520; PIDN:AA63215.1; PID:G187521
 C/Comment: This protein is a prominent component of the basement membrane that mediates cell-matrix interactions
 C/Genetics:
 A/Gene: GDB:LAMW2; LAMW
 A/Cross-references: GDB:132362; OMIM:156225
 A/Map position: 6Q22-6Q23
 C/Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
 C/Function: interact with cells and with other basement membrane proteins to promote cell-matrix interactions
 C/Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like I
 C/Keywords: basement membrane; calcium binding; coiled coil; extracellular matrix; Glyco
 F:161-107/Domain: laminin-type EGF-like homology #status atypical <LE02>
 F:110-165/Domain: laminin-type EGF-like homology <LE03>
 F:168-212/Domain: laminin-type EGF-like homology <LE04>
 F:527-567, 1071-1300/Region: 3DM and 2D9 binding
 F:811-972/Domain: laminin G repeat homology <LG1>
 F:1005-1165/Domain: laminin G repeat homology <LG2>
 F:1131-1254/Domain: laminin G repeat homology <LG3>
 F:1430-1578/Domain: laminin G repeat homology <LG4>
 F:1605-1751/Domain: laminin G repeat homology <LG5>
 F:120,238,255,341,451,542,557,561,658,669,686,767,881,1001,1076,1119,1192,1199,1289,1509

Matches 59; Conservative 23; Mismatches 67; Indels 13; Gaps 6;
 Qy 122 TIELEVRTASGLLLMQGVEGEAGQKDFSLGQDGHVFRYQLGSGEARLVSEDDPI 181
 |::|::|
 Db 1429 TIELEVRTEAESGLLFYM-ARINHA---DFATVQLRNGLFYFDLGSQDHTMIPTKI 1493
 |::|::|
 Qy 182 NDEGHWRTALRGRGSIQVDGEEELVSGRSPEFNA--VNAKGSVVIIGAP-DVATLIG 238
 |::|::|
 Db 1484 NDGQWHKIKIMRSKQEGILYVDG---ASNRTISPKKADILDVVGMLYVGGPLFIYTRRI 1540
 |::|::|
 Qy 239 GRFSSGITGCVKNLVLSHSARPQAPPPQLDLQHRQAQGANTR 280
 |::|::|
 Db 1541 GPVTSIDGCVRN--LHMAEAPADLEQPTSSSFHVGTCFANAQ 1580
 |::|::|

RESULT 15

A48216
 neurexin III-alpha secreted type 1 precursor - rat
 C/Species: Rattus norvegicus (Norway rat)
 C/Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Dec-2002
 C/Accession: A48216; B48216
 R;Ushkaryov, Y.A.; Suedhof, T.C.
 Proc. Natl. Acad. Sci. U.S.A. 90, 6410-6414, 1993
 A/Title: Neurexin IIIalpha: extensive alternative splicing generates membrane-bound and soluble isoforms
 A/Reference number: A48216; MUID:193342001; PMID:8341647
 A/Accession: A48216
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1438 <USH>
 A/Cross-references: GB:L14851
 A/Accession: B48216
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1368,1372-1438 <US2>
 A/Cross-references: GB:L14851
 C/Genetics:
 A/Introns: 1372/1
 C/Superfamily: neurexin; EGF homology
 C/Keywords: alternative splicing; brain; cell surface component; duplication; extracellu
 F:1-27/Domain: signal sequence #status predicted <SIG>
 F:202-234/Domain: EGF homology <EGF>
 F:651-663/Domain: EGF homology <EGF1>

Query Match 13.9%; Score 217; DB 2; Length 1438;
 Best Local Similarity 26.4%; Pred. No. 6.9e-08;
 Matches 67; Conservative 35; Mismatches 114; Indels 38; Gaps 8;

Qy 27 GFKGDLCE-----HEENPQREPCCLHGTC----QGRCLC-LPGF 63
 |::|::|
 Db 171 GFKGLMDLKYNSPRLGSGVLEAEGPCGER-PCENGGICPLLDGCHFTDCSTTGY 229
 |::|::|
 Qy 64 SGPRCQGGSHGIAESDWHLEGGNDAPGGVYFHDGDFLAFPHVFRSLSLSEVPETI 123
 |::|::|
 Db 230 GGTLCSEDEVSOGFLSHLMMSEQRKAREENVATFRGSEYLCY--DLSONPIQSSDEI 287
 |::|::|
 Qy 124 ELEVRTASGLLLMQGVEGEAGQKDFISLGLDGHVFRYQLGSGEARLVSEDDPI-- 181
 |::|::|
 Db 288 TLEPKTWRNGILL-----HTGSAIDVNLAKDGAIVLNLGSGAFRAIYE-PVNG 339
 |::|::|
 Qy 182 --NDGSHHWRTALRGRGSIQVDGEEELVSGRSPEFNAVNAKGSVVIIGAPDVATLIG 239
 |::|::|
 Db 340 KFNDNAHWDKVTRNLRQVTSVDGILTTTGTQEDYTMGLSDDDFFVYGGSPSTADLPFS 399
 |::|::|
 Qy 240 RFSSGITGCVKNLV 253
 |::|::|
 Db 400 PVSNNFMGLKEVV 413
 |::|::|

Search completed: March 9, 2004, 17:22:35
 Job time : 10.7114 secs

Query Match 14.3%; Score 223.5; DB 1; Length 1751;
 Best Local Similarity 36.4%; Pred. No. 2.9e-08;

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:19:08 ; Search time 20.844 Seconds
(without alignments)
2876.963 Million cell updates/sec

Title: US-10-006-011a-9
Perfect score: 1566
Sequence: 1 CERQCQHGATCMRAGEYEF.....QPLDLQHRAGANTRCPS 284

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/FCU_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/FCUUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Table with columns: Result No., Score, Query Match, Length, DB ID, Description. Contains 18 rows of search results.

Table with columns: 16, 333.5, 21.3, 1940, 13, US-10-016-283-34, Sequence 34, Appl. Contains 18 rows of alignment data.

ALIGNMENTS

RESULT 1
US-10-094-886-180
Sequence 180 Application US/10094886
Publication No. US20040002120A1
GENERAL INFORMATION:
APPLICANT: Kekuda, Ramesh
APPLICANT: Tchernev, Velizar T.
APPLICANT: Liu, Xiaohong
APPLICANT: Svytek, Kimberly A.
APPLICANT: Patturajan, Meeta
APPLICANT: Burgess, Catherine
APPLICANT: Vernet, Corine A.
APPLICANT: Li, Li
APPLICANT: Gorman, Linda
APPLICANT: Malyankar, Uriel M.
APPLICANT: Boldog, Ferenc
APPLICANT: Guo, Xiaojia
APPLICANT: Shenoy, Suresh
APPLICANT: Padigar, Muralidhara
APPLICANT: Taupier, Raymond J., Jr.
APPLICANT: Miller, Charles
APPLICANT: Casman, Stacie
APPLICANT: Pena, Carol
APPLICANT: Gangolli, Esha
APPLICANT: Gusev, Vladimir
APPLICANT: Smithson, Glennda
APPLICANT: Zerhusen, Bryan
APPLICANT: Gerlach, Valerie
APPLICANT: Pochart, Pascal
APPLICANT: Fernandes, Elma
APPLICANT: Shinkets, Richard
APPLICANT: Rastelli, Luca
APPLICANT: Spaderna, Steven
APPLICANT: LaRochelle, William
APPLICANT: Zhong, Mei
TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
FILE REFERENCE: 21402-290 B
CURRENT APPLICATION NUMBER: US/10/094,886

; CURRENT FILING DATE: 2002-03-07
 ; PRIOR APPLICATION NUMBER: 60/274,322
 ; PRIOR FILING DATE: 2001-03-08
 ; PRIOR APPLICATION NUMBER: 60/313,182
 ; PRIOR FILING DATE: 2001-08-17
 ; PRIOR APPLICATION NUMBER: 60/288,052
 ; PRIOR FILING DATE: 2001-05-02
 ; PRIOR APPLICATION NUMBER: 60/318,510
 ; PRIOR FILING DATE: 2001-09-10
 ; PRIOR APPLICATION NUMBER: 60/274,281
 ; PRIOR FILING DATE: 2001-03-08
 ; PRIOR APPLICATION NUMBER: 60/314,018
 ; PRIOR FILING DATE: 2001-08-21
 ; PRIOR APPLICATION NUMBER: 60/274,194
 ; PRIOR FILING DATE: 2001-03-08
 ; PRIOR APPLICATION NUMBER: 60/274,849
 ; PRIOR FILING DATE: 2001-03-09
 ; PRIOR APPLICATION NUMBER: 60/296,693
 ; PRIOR FILING DATE: 2001-06-07
 ; PRIOR APPLICATION NUMBER: 60/313,626
 ; PRIOR FILING DATE: 2001-08-21
 ; Remaining prior application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 238
 ; SOFTWARE: Patent in 2.1
 ; SEQ ID NO 180
 ; LENGTH: 1931
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-094-886-180

Query Match 22.8%; Score 357.5; DB 15; Length 1931;
 Best Local Similarity 35.4%; Pred. No. 2.3e-22;
 Matches 91; Conservative 29; Mismatches 110; Indels 27; Gaps 8;
 Qy 1 CERPCQGHGATCPAGEYEFQCLCRDGFKDLCEHEENPCQLREPCRLGGTCQ-----GT 55
 Db 1439 CLNPCHGGAPCNLEAGRPHCCPPRGVPTCADEKSPCQ-PNFCGGAAPCRVLPPEGGA 1497
 Qy 56 RCLCLPFGSPRQCGSGHGIAESDWHLESGGNDARGQYGFYFDDGFLAPPG-HVFSR 114
 Db 1498 QCECLPREGTFQCQTAS-----GQDGGSPFLADFNFGSHLELRGLHTFAR 1542
 Qy 115 SLPEVPTIELVRTSTASGLLWQGVYGEAGQGGKDFISLGLQDGHVFRYQLGSGEAR 174
 Db 1543 DLGE-KMALEWVFLARSPGSLLYNGQKTD--GKG-DFVSLALRRDRLEFRYDLGKGA 1598
 Qy 175 LVSEDPINDGEWHRVTALRGRGSIQVNDGEELVSGSPGNVAVNAKGSVIYGGAPDVA 234
 Db 1599 IRSREPTLGAWTRVSLERNRKGALRVDGDFRVLGSEVPVHTVLMKELPLVGGAPDFS 1658
 Qy 235 TLT-GGRFSSGITGVCK 250
 Db 1659 KLARAAVSSGFDGAIQ 1675

RESULT 2
 US-09-764-853-758
 ; Sequence 758, Application US/09764853
 ; Patent No. US20020090672A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: RIZ06
 ; CURRENT APPLICATION NUMBER: US/09/764,853
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 939
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 758
 ; LENGTH: 238
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:

; NAME/KEY: SITE
 ; LOCATION: (5)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-764-853-758
 Query Match 21.9%; Score 343; DB 9; Length 238;
 Best Local Similarity 33.5%; Pred. No. 3.1e-22;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
 Qy 34 EHEENPCQLREPCRLGGTC-----QCTRCLCLPFGSPRQCGSGHGIAESDWHLESGG 89
 Db 11 ENAAHPC-VRAFCAGGSCRPRKEGYDCDCPLGPEGLHCCOKAITEAI-----56
 Qy 90 DAP----GOYGAYFDDGFLAPPGHVFSRSLPEVPTIELVRTSTASGLLWQGVYGE 145
 Db 57 EIPQFIGRSYLYDNPDLKRVSG---SR-----NVFMRFKTTAKDGLLWRG---DSP 105
 Qy 146 AGGKDFISLGLQDGHVFRYQLGSGEARLYSEDPINDGEWHRVTALRGRGSIQVNDGE 205
 Db 106 MRPNDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 165
 Qy 206 ELVSGRSPGNVAVNAKGSVIYGGAPDVAITLGGFRSSGITGVCKNLVLSHARFGAPPQ 265
 Db 166 GARTGKSPGMFQNLINGALYVGGMKEIALTHTNRQYMRGLVGCISHFTLST-----DY 218
 Qy 266 PLDLQHRAQAQANTRPC 282
 Db 219 HILSLVEDAVDGNINTC 235

RESULT 3
 US-09-764-898-252
 ; Sequence 252, Application US/09764898
 ; Patent No. US20020090673A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: RJZ01
 ; CURRENT APPLICATION NUMBER: US/09/764,898
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 311
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 252
 ; LENGTH: 238
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (5)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-764-898-252

Query Match 21.9%; Score 343; DB 9; Length 238;
 Best Local Similarity 33.5%; Pred. No. 3.1e-22;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
 Qy 34 EHEENPCQLREPCRLGGTC-----QCTRCLCLPFGSPRQCGSGHGIAESDWHLESGG 89
 Db 11 ENAAHPC-VRAFCAGGSCRPRKEGYDCDCPLGPEGLHCCOKAITEAI-----56
 Qy 90 DAP----GOYGAYFDDGFLAPPGHVFSRSLPEVPTIELVRTSTASGLLWQGVYGE 145
 Db 57 EIPQFIGRSYLYDNPDLKRVSG---SR-----NVFMRFKTTAKDGLLWRG---DSP 105
 Qy 146 AGGKDFISLGLQDGHVFRYQLGSGEARLYSEDPINDGEWHRVTALRGRGSIQVNDGE 205
 Db 106 MRPNDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDGQSGKITVDDY 165
 Qy 206 ELVSGRSPGNVAVNAKGSVIYGGAPDVAITLGGFRSSGITGVCKNLVLSHARFGAPPQ 265
 Db 166 GARTGKSPGMFQNLINGALYVGGMKEIALTHTNRQYMRGLVGCISHFTLST-----DY 218

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

Query Match 21.9%; Score 343; DB 14; Length 238;
Best Local Similarity 33.5%; Pred No. 3.1e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

Query Match 21.9%; Score 343; DB 10; Length 238;
Best Local Similarity 33.5%; Pred No. 3.1e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

RESULT 4
US-09-764-881-158
; Sequence 158, Application US/09764881

RESULT 5
US-10-073-865-121
; Sequence 121, Application US/10073865

FILE REFERENCE: PT207C1
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
CURRENT APPLICATION NUMBER: US/10/242,747
CURRENT FILING DATE: 2002-09-13
PRIOR APPLICATION NUMBER: 09/764,881
PRIOR FILING DATE: 2001-01-17

FILE REFERENCE: PT207C1
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
CURRENT APPLICATION NUMBER: US/10/242,747
CURRENT FILING DATE: 2002-09-13
PRIOR APPLICATION NUMBER: 09/764,881
PRIOR FILING DATE: 2001-01-17

US-10-073-865-121
; Sequence 121, Application US/10073865
; Publication No. US20030044904A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.

US-10-073-865-121
; Sequence 121, Application US/10073865
; Publication No. US20030044904A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.

Query Match 21.9%; Score 343; DB 15; Length 238;
Best Local Similarity 33.5%; Pred. No. 3.1e-22;

Query Match 21.9%; Score 343; DB 10; Length 238;
Best Local Similarity 33.5%; Pred. No. 3.1e-22;

Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
 QY 34 EHEENPCOLREPLHGTC---QTRCLCLPFGSPPRCQSGHGIABSDWHLGSGGN 89
 Db 11 ENAAHPC-VRAPCAHGSCRRKREGYDCDPLGPEGLHCQKAIIEAI-----56
 QY 90 DAP---GOYGAYFHDDGFLAFPGHVFVSRSLPEVPEVETIEVETSTASGILLMQGVEVGE 145
 Db 57 EIPQFIGRSYLYDNPDLKRVSG---SRS-----NVFMFKTKAKDGLLWRG--DSP 105
 QY 146 AGGKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGHWHTALREGRRSIQVDGE 205
 Db 106 MRPNSEDFISLGLRDGALVFSYVLSGVSASIMVNGSFNDRWHRVKAVRDQSGSKITVDDY 165
 QY 206 ELVSGRSPGNVAVNAKGSYVIGGAPDVATLTGGRFSSGITGCYKLVLSHARPGAPPQ 265
 Db 166 GARTGKSPGMQRLNIALYVGMKEIALHTNQYMRGLVGCISHFTLST-----DY 218
 QY 266 PLDLOHRAQAGANTRPC 282
 Db 219 HISLVEDAVDGKNINTC 235

RESULT 7
 US-09-978-249-10
 ; Sequence 10, Application US/09978249
 ; Patent No. US20020106780A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fiscella, et al.
 ; TITLE OF INVENTION: Extracellular Matrix Polynucleotides, Polypeptides, and Antibodies
 ; FILE REFERENCE: PT054PI
 ; CURRENT APPLICATION NUMBER: US/09/978,249
 ; CURRENT FILING DATE: 2001-10-17
 ; PRIOR APPLICATION NUMBER: PCT/US01/11643
 ; PRIOR FILING DATE: 2001-04-11
 ; PRIOR APPLICATION NUMBER: 60/198,123
 ; PRIOR FILING DATE: 2000-04-18
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 10
 ; LENGTH: 375
 ; TYPE: PRP
 ; ORGANISM: Homo sapiens
 US-09-978-249-10

Query Match 21.9%; Score 343; DB 9; Length 375;
 Best Local Similarity 33.5%; Pred. No. 5.5e-22;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
 QY 34 EHEENPCOLREPLHGTC---QTRCLCLPFGSPPRCQSGHGIABSDWHLGSGGN 89
 Db 148 ENAAHPC-VRAPCAHGSCRRKREGYDCDPLGPEGLHCQKAIIEAI-----193
 QY 90 DAP---GOYGAYFHDDGFLAFPGHVFVSRSLPEVPEVETIEVETSTASGILLMQGVEVGE 145
 Db 194 EIPQFIGRSYLYDNPDLKRVSG---SRS-----NVFMFKTKAKDGLLWRG--DSP 242
 QY 146 AGGKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGHWHTALREGRRSIQVDGE 205
 Db 243 MRPNSEDFISLGLRDGALVFSYVLSGVSASIMVNGSFNDRWHRVKAVRDQSGSKITVDDY 302
 QY 206 ELVSGRSPGNVAVNAKGSYVIGGAPDVATLTGGRFSSGITGCYKLVLSHARPGAPPQ 265
 Db 303 GARTGKSPGMQRLNIALYVGMKEIALHTNQYMRGLVGCISHFTLST-----DY 355
 QY 266 PLDLOHRAQAGANTRPC 282
 Db 356 HISLVEDAVDGKNINTC 372

RESULT 8
 US-09-764-853-541
 ; Sequence 541, Application US/09764853

Patent No. US20020090672A1
 GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 FILE REFERENCE: PJ206
 CURRENT APPLICATION NUMBER: US/09/764,853
 CURRENT FILING DATE: 2001-01-17
 Prior application data removed - consult PALM or file wrapper
 NUMBER OF SEQ ID NOS: 939
 SOFTWARE: Patent In Ver. 2.0
 SEQ ID NO 541
 LENGTH: 432
 TYPE: PRP
 ORGANISM: Homo sapiens
 US-09-764-853-541
 Query Match 21.9%; Score 343; DB 9; Length 432;
 Best Local Similarity 33.5%; Pred. No. 6.6e-22;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
 QY 34 EHEENPCOLREPLHGTC---QTRCLCLPFGSPPRCQSGHGIABSDWHLGSGGN 89
 Db 205 ENAAHPC-VRAPCAHGSCRRKREGYDCDPLGPEGLHCQKAIIEAI-----250
 QY 90 DAP---GOYGAYFHDDGFLAFPGHVFVSRSLPEVPEVETIEVETSTASGILLMQGVEVGE 145
 Db 251 EIPQFIGRSYLYDNPDLKRVSG---SRS-----NVFMFKTKAKDGLLWRG--DSP 299
 QY 146 AGGKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGHWHTALREGRRSIQVDGE 205
 Db 300 MRPNSEDFISLGLRDGALVFSYVLSGVSASIMVNGSFNDRWHRVKAVRDQSGSKITVDDY 359
 QY 206 ELVSGRSPGNVAVNAKGSYVIGGAPDVATLTGGRFSSGITGCYKLVLSHARPGAPPQ 265
 Db 360 GARTGKSPGMQRLNIALYVGMKEIALHTNQYMRGLVGCISHFTLST-----DY 412
 QY 266 PLDLOHRAQAGANTRPC 282
 Db 413 HISLVEDAVDGKNINTC 429

RESULT 9
 US-09-764-898-179
 ; Sequence 179, Application US/09764898
 ; Patent No. US20020090673A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PJ201
 ; CURRENT APPLICATION NUMBER: US/09/764,898
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 311
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 179
 ; LENGTH: 432
 ; TYPE: PRP
 ; ORGANISM: Homo sapiens
 US-09-764-898-179

Query Match 21.9%; Score 343; DB 9; Length 432;
 Best Local Similarity 33.5%; Pred. No. 6.6e-22;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
 QY 34 EHEENPCOLREPLHGTC---QTRCLCLPFGSPPRCQSGHGIABSDWHLGSGGN 89
 Db 205 ENAAHPC-VRAPCAHGSCRRKREGYDCDPLGPEGLHCQKAIIEAI-----250
 QY 90 DAP---GOYGAYFHDDGFLAFPGHVFVSRSLPEVPEVETIEVETSTASGILLMQGVEVGE 145
 Db 251 EIPQFIGRSYLYDNPDLKRVSG---SRS-----NVFMFKTKAKDGLLWRG--DSP 299
 QY 146 AGGKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGHWHTALREGRRSIQVDGE 205

```

Db 300 MRPNDFISLGRDCLVALFVSNLGSVAVSIVVNGSFNDRWHRVAVKAVRVDGSGKTIIVDDY 359
QY 206 ELVSRSPGPNVAVNAKSVVIGGAPDVATLTGGRFSSGTTGCVKNLVLHSARFGAPPQ 265
Db 360 GARTKSPGMQMLNGALVYGGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 412
QY 266 PLDLOHRAQAQANTRPC 282
Db 413 HISLVEDAVDGKNINTC 429

```

```

RESULT 10
US-09-764-881-102
; Sequence 102, Application US/09764881
; Publication No. US20030125246A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ07
; CURRENT APPLICATION NUMBER: US/09/764,881
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 102
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-881-102

```

```

Query Match 21.9%; Score 343; DB 10; Length 432;
Best Local Similarity 33.5%; Pred. No. 6.6e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

QY 34 EHEENPCOLREPCLEHGTC-----QGTCLCLPFGSPPRCQOQSGHGIABSDWHLESGGN 89
Db 205 ENAAHPC-VRAPCAHGSCRPKEGYDCDCPLGFEGLHCOKAIEAI----- 250
QY 90 DAP-----GOYGAIFHDDGFLAPGHVFRSRLPEVPEIIEVTRTSTASGLLLWQGVVEGE 145
Db 251 EIPOFIGRSYLYDNPDLKRVSG---SRS-----NVFMRFKTTAKDGLLLWRG--DSP 299
QY 146 AQGGKDFISLGLQDGLHVFYRQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
Db 300 MRPNDFISLGRDCLVALFVSNLGSVAVSIVVNGSFNDRWHRVAVKAVRVDGSGKTIIVDDY 359
QY 206 ELVSRSPGPNVAVNAKSVVIGGAPDVATLTGGRFSSGTTGCVKNLVLHSARFGAPPQ 265
Db 360 GARTKSPGMQMLNGALVYGGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 412
QY 266 PLDLOHRAQAQANTRPC 282
Db 413 HISLVEDAVDGKNINTC 429

```

```

RESULT 11
US-10-073-865-78
; Sequence 78, Application US/10073865
; Publication No. US20030044904A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZ09CI
; CURRENT APPLICATION NUMBER: US/10/073,865
; CURRENT FILING DATE: 2002-02-14
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 154
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 78
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens

```

```

US-10-073-865-78
Query Match 21.9%; Score 343; DB 14; Length 432;
Best Local Similarity 33.5%; Pred. No. 6.6e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

QY 34 EHEENPCOLREPCLEHGTC-----QGTCLCLPFGSPPRCQOQSGHGIABSDWHLESGGN 89
Db 205 ENAAHPC-VRAPCAHGSCRPKEGYDCDCPLGFEGLHCOKAIEAI----- 250
QY 90 DAP-----GOYGAIFHDDGFLAPGHVFRSRLPEVPEIIEVTRTSTASGLLLWQGVVEGE 145
Db 251 EIPOFIGRSYLYDNPDLKRVSG---SRS-----NVFMRFKTTAKDGLLLWRG--DSP 299
QY 146 AQGGKDFISLGLQDGLHVFYRQLGSGEARLVSEDPINDGEWHRVTALREGRRGSIQVDGE 205
Db 300 MRPNDFISLGRDCLVALFVSNLGSVAVSIVVNGSFNDRWHRVAVKAVRVDGSGKTIIVDDY 359
QY 206 ELVSRSPGPNVAVNAKSVVIGGAPDVATLTGGRFSSGTTGCVKNLVLHSARFGAPPQ 265
Db 360 GARTKSPGMQMLNGALVYGGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 412
QY 266 PLDLOHRAQAQANTRPC 282
Db 413 HISLVEDAVDGKNINTC 429

```

```

RESULT 12
US-10-242-747-102
; Sequence 102, Application US/10242747
; Publication No. US20040005577A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ07C1
; CURRENT APPLICATION NUMBER: US/10/242,747
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,881
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 102
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-242-747-102

```

```

Query Match 21.9%; Score 343; DB 15; Length 432;
Best Local Similarity 33.5%; Pred. No. 6.6e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

QY 34 EHEENPCOLREPCLEHGTC-----QGTCLCLPFGSPPRCQOQSGHGIABSDWHLESGGN 89
Db 205 ENAAHPC-VRAPCAHGSCRPKEGYDCDCPLGFEGLHCOKAIEAI----- 250

```

```

RESULT 13
US-10-073-865-78
; Sequence 78, Application US/10073865
; Publication No. US20030044904A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZ09CI
; CURRENT APPLICATION NUMBER: US/10/073,865
; CURRENT FILING DATE: 2002-02-14
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 154
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 78
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo sapiens

```

```

Qy 90 DAP----GOYGAYFDDGFLAFPGHVSRLPEVETIELEVRTSTAGSLLLWQVEVGE 145
Db 251 EIQFIGRSYLTVDNPDILKRVSG---SR-----NVFMRFKTTAKDGLLLWRG--DSP 299
Qy 146 AGQKDFISLGLQDHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRSIQVDGE 205
Db 300 MRNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDQSGSKITVDDY 359
Qy 206 ELVSGRSPGNVAVNAKGSVIYGGAPDVALTGGFRFSSGITGVKNLVLHSAARPCAPPQ 265
Db 360 GARTGKSPGMRQLNGLVGGMKEIALHTNRQVMRGLVGCISHFTLST-----DY 412
Qy 266 PLDQHRQAQANTRPC 282
Db 413 HISLVEDAVDGNKNTC 429

```

```

RESULT 13
US-10-104-047-3058
; Sequence 3058, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cdna
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3058
; LENGTH: 463
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-104-047-3058

```

```

Query Match 21.9%; Score 343; DB 15; Length 463;
Best Local Similarity 33.5%; Pred. No. 7.2e-22;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

Qy 34 EHENPCQLREPLHGTC----QGTFCICLFGSPRCQCGSGHGAIESDWHLEGGGN 89
Db 236 ENAAHPC-VRAPCAHGSCPRKEGYDCCPLGFEGLHCQKAIIEAI----- 281
Qy 90 DAP----GOYGAYFDDGFLAFPGHVSRLPEVETIELEVRTSTAGSLLLWQVEVGE 145
Db 282 EIQFIGRSYLTVDNPDILKRVSG---SR-----NVFMRFKTTAKDGLLLWRG--DSP 330
Qy 146 AGQKDFISLGLQDHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRSIQVDGE 205
Db 331 MRNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDQSGSKITVDDY 390
Qy 206 ELVSGRSPGNVAVNAKGSVIYGGAPDVALTGGFRFSSGITGVKNLVLHSAARPCAPPQ 265
Db 391 GARTGKSPGMRQLNGLVGGMKEIALHTNRQVMRGLVGCISHFTLST-----DY 443
Qy 266 PLDQHRQAQANTRPC 282
Db 444 HISLVEDAVDGNKNTC 460

```

```

RESULT 14
US-10-108-260A-4433
; Sequence 4433, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cdna
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27

```

```

; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4433
; LENGTH: 775
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-108-260A-4433

Query Match 21.9%; Score 343; DB 15; Length 775;
Best Local Similarity 33.5%; Pred. No. 1.4e-21;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

Qy 34 EHENPCQLREPLHGTC----QGTFCICLFGSPRCQCGSGHGAIESDWHLEGGGN 89
Db 548 ENAAHPC-VRAPCAHGSCPRKEGYDCCPLGFEGLHCQKAIIEAI----- 593
Qy 90 DAP----GOYGAYFDDGFLAFPGHVSRLPEVETIELEVRTSTAGSLLLWQVEVGE 145
Db 594 EIQFIGRSYLTVDNPDILKRVSG---SR-----NVFMRFKTTAKDGLLLWRG--DSP 642
Qy 146 AGQKDFISLGLQDHLVFRYQLGSGEARLVSEDPINDGEWHRVTALREGRRSIQVDGE 205
Db 643 MRNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDQSGSKITVDDY 702
Qy 206 ELVSGRSPGNVAVNAKGSVIYGGAPDVALTGGFRFSSGITGVKNLVLHSAARPCAPPQ 265
Db 703 GARTGKSPGMRQLNGLVGGMKEIALHTNRQVMRGLVGCISHFTLST-----DY 755
Qy 266 PLDQHRQAQANTRPC 282
Db 756 HISLVEDAVDGNKNTC 772

```

```

RESULT 15
US-10-264-049-2933
; Sequence 2933, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Bires et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133PI
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; NUMBER OF SEQ ID NOS: 4380
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2933
; LENGTH: 406
; TYPE: PR1
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (82)_
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (95)_
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (96)_
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2933

```

```

Query Match 21.7%; Score 340.5; DB 15; Length 406;
Best Local Similarity 29.2%; Pred. No. 1e-21;
Matches 99; Conservative 38; Mismatches 113; Indels 89; Gaps 11;

Qy 1 CERQPCQHGATCMP-----AGEYEFQCLCR-----DFEKG----- 30

```

Thu Mar 11 09:25:23 2004

```

Db 101 CGRRPRVLGEGSPVHTVNLKELPLVVGAPDFSKLARAANAIVSSGFDGAIQLVSLGGRQLL 160
QY 31 -----DLCEHEENFCOLR--EPCILHGGTCOGTR----CLCLPGFSGPRCQOQSGHG 75
Db 161 TPEHVLROVDVTSFAGHPCTRASGHPCLNGASCVPREAAAYVCLCPGGFSGPHCEKGL--- 217
QY 76 IABSDWHLEGGCGNDAPGOYGAFFHDDGFLAAPPGHVFSRSLPEVPEI-----IELEV 127
Db 218 -----VEKSAG-----DVTLAPDGRFTVEYLNATYSEKALQSNHFELS 258
QY 128 RTSTASGLLLMOQVEVGBAGQKDFISLQDGHILVFRYQLGSGEARLIVSEDPINDGEWH 187
Db 259 RTEATQGLVMS-----GRATERADYVALAIVDGHILQLSYNLGSPVLRSTVPVNTNRWL 314
QY 188 RVTALREGRRSIQVDGEBELVSRGFGFNVAIVKGSVYICGAPD--VATLTGGRFSSGI 245
Db 315 RYVAHREQREGSLQVNEAPVTGSSPLGATQDITDGLMGLGELPELVGPALPKAYGTGF 374
QY 246 TGCVKNLVLSARPGAPPQPLDQHRQAQAGANTRPCPS 284
Db 375 VGCLRDVVVGR-----HPLHLEDAVTKPELRPCPT 405

```

Search completed: March 9, 2004, 17:25:15
Job time : 21.844 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:16:23 ; Search time 11.1326 Seconds
(without alignments)
1317.011 Million cell updates/sec

Title: US-10-006-011A-9
Perfect score: 1566
Sequence: 1 CERPCQHGATCMPAGEYEF.....QPLDLQHRAGAGNFRPCFS 284

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	333.5	21.3	1940	2	US-08-644-271-30
2	333.5	21.3	1940	4	US-09-077-955-34
3	330	21.1	294	4	US-09-077-955-29
4	330	21.1	338	4	US-09-077-955-28
5	330	21.1	390	4	US-09-077-955-27
6	330	21.1	440	4	US-09-077-955-26
7	330	21.1	456	4	US-09-077-955-25
8	330	21.1	492	2	US-08-644-271-32
9	330	21.1	492	4	US-09-077-955-36
10	326	20.8	256	4	US-09-077-955-30
11	261	16.7	216	4	US-09-077-955-31
12	253.5	16.2	1525	3	US-09-191-647-2
13	253.5	16.2	1525	3	US-09-540-245A-2
14	253.5	16.2	1525	3	US-09-540-153-2
15	249	15.9	735	3	US-09-191-647-9
16	249	15.9	735	3	US-09-540-245A-9
17	249	15.9	735	3	US-09-540-153-9
18	245.5	15.7	1529	4	US-09-312-283C-396
19	239	15.3	1523	4	US-09-182-024A-2
20	226.5	14.5	716	4	US-09-312-283C-183
21	226.5	14.5	771	3	US-09-188-930-183
22	223.5	14.3	1130	2	US-08-460-309-2
23	223.5	14.3	1130	2	US-08-125-077-2
24	223.5	14.3	1130	6	544158-2
25	223.5	14.3	3088	4	US-09-562-702A-8
26	223.5	14.3	3089	4	US-09-562-702A-4
27	223.5	14.3	3110	4	US-09-562-702A-2

Sequence 6, Appli
Sequence 7, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 12, Appli
Sequence 10, Appli
Sequence 2, Appli
Sequence 4, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 7, Appli
Sequence 7, Appli
Sequence 2, Appli

28 223.5 14.3 3110 4 US-09-562-702A-6
29 223.5 14.3 3110 4 US-09-561-709B-7
30 223.5 14.3 3111 2 US-08-460-309-4
31 223.5 14.3 3111 2 US-08-125-077-4
32 207.5 13.3 3084 4 US-09-562-702A-12
33 207.5 13.3 3106 4 US-09-562-702A-10
34 206 13.2 3571 4 US-09-911-842A-2
35 202.5 12.9 3594 4 US-09-911-842A-4
36 193.5 12.7 160 3 US-09-191-647-5
37 193.5 12.7 160 3 US-09-540-245A-5
38 193.5 12.7 160 3 US-09-540-153-5
39 199 12.7 2321 4 US-09-230-652-2
40 188.5 12.0 3075 2 US-08-460-309-5
41 188.5 12.0 3075 2 US-08-125-077-5
42 183 11.7 1480 3 US-09-191-647-7
43 183 11.7 1480 3 US-09-540-245A-7
44 183 11.7 1480 3 US-09-540-153-7
45 183 11.7 1480 5 PCT-US91-09055-2

ALIGNMENTS

RESULT 1
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; TITLE OF INVENTION: AND LIGANDS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/644,271
; FILING DATE: 10-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 60/008,657
; FILING DATE: 15-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Covert, Robert J
; REGISTRATION NUMBER: 36,108
; REFERENCE/DOCKET NUMBER: REG 195A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
; TELEX:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Rat Agrin
; LOCATION: 1..1940
; OTHER INFORMATION:
; US-08-644-271-30

Query Match 21.3%; Score 333.5; DB 2; Length 1940;

Best Local Similarity 36.0%; Pred. No. 9.2e-22;
 Matches 94; Conservative 28; Mismatches 108; Indels 31; Gaps 10;

QY 1 CERQPCOHGATCMPAGEYFQCLCRDGFKGLDCEHEENPCQLREPCLHGTCQ-----GT 55
 Db 1444 CLPNCHGGALCOALEAGMFLCQCPRGPTCADEKSPCO-PNFCGAAFCRVLSSGGA 1502

QY 56 RCLCLFPGFSPRCQQSGHGAESDWHLEGGSDNAPQGYGAYFHDDGFLAAPP-GVFSR 114
 Db 1503 KCECPLGRSGTFCQT-----VLETAGSR-----PFLADFNFGFSYLELKLHTFER 1547

QY 115 SLPEVPELLEVRTSTAGSLLLMWQVVEGAGGCKDFISLQDGHVFRYQLGSGEAR 174
 Db 1548 DLGE-KWALEMVLARPGSGLLYNGQKTD--GKG-DFVSLAHNRHLEFCYDLGKGAAV 1603

QY 175 LVSEDPINDGEWHRVTALEGRRGSIQVDGEEELVSRGSPG-----PNVAVNAKGSVYIGGA 230
 Db 1604 IRSKEPALTGWVFLERNRGRKALQVGDGPRVLGSEPKRKPVHTMLNLKPELYIGGA 1663

QY 231 PDVATLTLT-GGRFSSGITGCVK 250
 Db 1664 PPFSKLARGAANVSSGFSGVIIQ 1684

RESULT 2
 US-09-077-955-34
 ; Sequence 34, Application US/09077955A
 ; Patent No. 6413740
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela et al., David M.
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
 ; FILE REFERENCE: REG195-B-PCT-US
 ; CURRENT APPLICATION NUMBER: US/09/077,955A
 ; EARLIER FILING DATE: 1998-09-10
 ; EARLIER APPLICATION NUMBER: PCT/US96/20696
 ; EARLIER FILING DATE: 1996-12-13
 ; EARLIER APPLICATION NUMBER: 08/644,271
 ; EARLIER FILING DATE: 1996-05-10
 ; EARLIER APPLICATION NUMBER: 60/008,657
 ; EARLIER FILING DATE: 1995-12-15
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 34
 ; LENGTH: 1940
 ; TYPE: PRT
 ; ORGANISM: Rattus sp.
 US-09-077-955-34

Query Match 21.3%; Score 333.5; DB 4; Length 1940;
 Best Local Similarity 36.0%; Pred. No. 9.2e-22;
 Matches 94; Conservative 28; Mismatches 108; Indels 31; Gaps 10;

QY 1 CERQPCOHGATCMPAGEYFQCLCRDGFKGLDCEHEENPCQLREPCLHGTCQ-----GT 55
 Db 1444 CLPNCHGGALCOALEAGMFLCQCPRGPTCADEKSPCO-PNFCGAAFCRVLSSGGA 1502

QY 56 RCLCLFPGFSPRCQQSGHGAESDWHLEGGSDNAPQGYGAYFHDDGFLAAPP-GVFSR 114
 Db 1503 KCECPLGRSGTFCQT-----VLETAGSR-----PFLADFNFGFSYLELKLHTFER 1547

QY 115 SLPEVPELLEVRTSTAGSLLLMWQVVEGAGGCKDFISLQDGHVFRYQLGSGEAR 174
 Db 1548 DLGE-KWALEMVLARPGSGLLYNGQKTD--GKG-DFVSLAHNRHLEFCYDLGKGAAV 1603

QY 175 LVSEDPINDGEWHRVTALEGRRGSIQVDGEEELVSRGSPG-----PNVAVNAKGSVYIGGA 230
 Db 1604 IRSKEPALTGWVFLERNRGRKALQVGDGPRVLGSEPKRKPVHTMLNLKPELYIGGA 1663

QY 231 PDVATLTLT-GGRFSSGITGCVK 250
 Db 1664 PPFSKLARGAANVSSGFSGVIIQ 1684

RESULT 3
 US-09-077-955-29
 ; Sequence 29, Application US/09077955A
 ; Patent No. 6413740
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela et al., David M.
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
 ; FILE REFERENCE: REG195-B-PCT-US
 ; CURRENT APPLICATION NUMBER: US/09/077,955A
 ; EARLIER FILING DATE: 1998-09-10
 ; EARLIER APPLICATION NUMBER: PCT/US96/20696
 ; EARLIER FILING DATE: 1996-12-13
 ; EARLIER APPLICATION NUMBER: 08/644,271
 ; EARLIER FILING DATE: 1996-05-10
 ; EARLIER APPLICATION NUMBER: 60/008,657
 ; EARLIER FILING DATE: 1995-12-15
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 29
 ; LENGTH: 294
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-077-955-29

Query Match 21.1%; Score 330; DB 4; Length 294;
 Best Local Similarity 31.3%; Pred. No. 1.5e-22;
 Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHEENPCQLR--EPCLHGTCQGTTR----CLCLPFGFSPRCQQSGHGAESDWHLE 84
 Db 50 DVTSFAGHPCTFASGHFLCNGASCVPREAAVYVCLCFGGFSGPHCEKGL-----VE 99

QY 85 GGGNDAPQGYGAYFHDDGFLAAPPCHVFSRSLPEVPEP-----ILEVLR 128
 Db 100 KSAQ-----DVTDLAFDGRTFVYLAANVTESELANEIPVEKALQSNHFFELSLR 147

QY 129 TSTAGSLLLMWQVVEGAGGCKDFISLQDGHVFRYQLGSGEARLIVSDPINDGEWHR 188
 Db 148 TEATQGLVMS---GRATERADVALAIVDGHQLSYNLGSPVVLVLRSTVPTVNTRWLR 203

QY 189 VTALREGRRSIQVDGEEELVSRGSPGPNVAVNAKGSVYIGGAPD--VAITLGGFRFSSGIT 246
 Db 204 VVAHREQREGSLQVNEAPVTGSSPLGATQLDITDGDALMLGLGPELPGVPALPKAYGTGFV 263

QY 247 GCVKNLVLSARPGAPPPQPLDLQHRAQAGANTRPCPS 284
 Db 264 GCLRDVVVGR-----HPLHLEDAVTKPELRPCPT 293

RESULT 4
 US-09-077-955-28
 ; Sequence 28, Application US/09077955A
 ; Patent No. 6413740
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela et al., David M.
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
 ; FILE REFERENCE: REG195-B-PCT-US
 ; CURRENT APPLICATION NUMBER: US/09/077,955A
 ; EARLIER FILING DATE: 1998-09-10
 ; EARLIER APPLICATION NUMBER: PCT/US96/20696
 ; EARLIER FILING DATE: 1996-12-13
 ; EARLIER APPLICATION NUMBER: 08/644,271
 ; EARLIER FILING DATE: 1996-05-10
 ; EARLIER APPLICATION NUMBER: 60/008,657
 ; EARLIER FILING DATE: 1995-12-15
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 28
 ; LENGTH: 338
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-077-955-28

Query Match 21.3%; Score 333.5; DB 4; Length 1940;
 Best Local Similarity 36.0%; Pred. No. 9.2e-22;
 Matches 94; Conservative 28; Mismatches 108; Indels 31; Gaps 10;

QY 1 CERQPCOHGATCMPAGEYFQCLCRDGFKGLDCEHEENPCQLREPCLHGTCQ-----GT 55
 Db 1444 CLPNCHGGALCOALEAGMFLCQCPRGPTCADEKSPCO-PNFCGAAFCRVLSSGGA 1502

QY 56 RCLCLFPGFSPRCQQSGHGAESDWHLEGGSDNAPQGYGAYFHDDGFLAAPP-GVFSR 114
 Db 1503 KCECPLGRSGTFCQT-----VLETAGSR-----PFLADFNFGFSYLELKLHTFER 1547

QY 115 SLPEVPELLEVRTSTAGSLLLMWQVVEGAGGCKDFISLQDGHVFRYQLGSGEAR 174
 Db 1548 DLGE-KWALEMVLARPGSGLLYNGQKTD--GKG-DFVSLAHNRHLEFCYDLGKGAAV 1603

QY 175 LVSEDPINDGEWHRVTALEGRRGSIQVDGEEELVSRGSPG-----PNVAVNAKGSVYIGGA 230
 Db 1604 IRSKEPALTGWVFLERNRGRKALQVGDGPRVLGSEPKRKPVHTMLNLKPELYIGGA 1663

QY 231 PDVATLTLT-GGRFSSGITGCVK 250
 Db 1664 PPFSKLARGAANVSSGFSGVIIQ 1684

```

Query Match          21.1%; Score 330; DB 4; Length 338;
Best Local Similarity 31.3%; Pred. No. 1.9e-22;
Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPCFHGTCQGTR----CLCLPFGSGPRCCQSGHGIAESDWHLE 84
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
94 DVTSFAGHPCTRASGHPCLNGASCVPREAAVYVCLCPGFSGPHCEKGL-----VE 143
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 85 GSGNDAPGQYGFHDDGFLAFPGHVFYRSLPEVPT-----IELEVR 128
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 144 KSAG-----DVDTLAFDGRFVYELNVAVTESELANEIPVEKALQSNHFEISLR 191
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 129 TSTASGLLLWQVGEAGQKDISLQDGHVFRYQLGSGEARLVSDEDPINDGEWHR 188
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db 192 TEATQGLVWS----GKATERADYVALAIVDGHQLSINLGSQPVVLRSTVAVTNRWLR 247
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
QY 189 VTALREGRGSIQVDBELVSRSPGNVAVNAKGSVYIGAPD--VAITLGGRFSSGIT 246
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db 248 VVAHREQREGSLQVNEAPVTGSSPLGATQDLDGALWLGGLPELFPVGPALPKAYGTGFV 307
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
QY 247 GCVKNLVLSHARPGAPPDLPDQHRAGANTRPCPS 284
   |||:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db 308 GCLRDVVVGR-----HPLHLEDAVTKPELRPCPT 337
   |||:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:

```

```

RESULT 5
US-09-077-955-27
; Sequence 27, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 27
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-27

```

```

Query Match          21.1%; Score 330; DB 4; Length 390;
Best Local Similarity 31.3%; Pred. No. 2.3e-22;
Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPCFHGTCQGTR----CLCLPFGSGPRCCQSGHGIAESDWHLE 84
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db 146 DVTSFAGHPCTRASGHPCLNGASCVPREAAVYVCLCPGFSGPHCEKGL-----VE 195
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
QY 85 GSGNDAPGQYGFHDDGFLAFPGHVFYRSLPEVPT-----IELEVR 128
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db 196 KSAG-----DVDTLAFDGRFVYELNVAVTESELANEIPVEKALQSNHFEISLR 243
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
QY 129 TSTASGLLLWQVGEAGQKDISLQDGHVFRYQLGSGEARLVSDEDPINDGEWHR 188
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db 244 TEATQGLVWS----GKATERADYVALAIVDGHQLSINLGSQPVVLRSTVAVTNRWLR 299
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
QY 189 VTALREGRGSIQVDBELVSRSPGNVAVNAKGSVYIGAPD--VAITLGGRFSSGIT 246
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db 300 VVAHREQREGSLQVNEAPVTGSSPLGATQDLDGALWLGGLPELFPVGPALPKAYGTGFV 359
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
QY 247 GCVKNLVLSHARPGAPPDLPDQHRAGANTRPCPS 284
   |||:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db 360 GCLRDVVVGR-----HPLHLEDAVTKPELRPCPT 389
   |||:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:

```

```

RESULT 6
US-09-077-955-26
; Sequence 26, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 26
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-26

```

```

Query Match          21.1%; Score 330; DB 4; Length 440;
Best Local Similarity 31.3%; Pred. No. 2.7e-22;
Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPCFHGTCQGTR----CLCLPFGSGPRCCQSGHGIAESDWHLE 84
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db 196 DVTSFAGHPCTRASGHPCLNGASCVPREAAVYVCLCPGFSGPHCEKGL-----VE 245
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
QY 85 GSGNDAPGQYGFHDDGFLAFPGHVFYRSLPEVPT-----IELEVR 128
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db 246 KSAG-----DVDTLAFDGRFVYELNVAVTESELANEIPVEKALQSNHFEISLR 293
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
QY 129 TSTASGLLLWQVGEAGQKDISLQDGHVFRYQLGSGEARLVSDEDPINDGEWHR 188
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db 294 TEATQGLVWS----GKATERADYVALAIVDGHQLSINLGSQPVVLRSTVAVTNRWLR 349
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
QY 189 VTALREGRGSIQVDBELVSRSPGNVAVNAKGSVYIGAPD--VAITLGGRFSSGIT 246
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db 350 VVAHREQREGSLQVNEAPVTGSSPLGATQDLDGALWLGGLPELFPVGPALPKAYGTGFV 409
   |||:|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
QY 247 GCVKNLVLSHARPGAPPDLPDQHRAGANTRPCPS 284
   |||:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db 410 GCLRDVVVGR-----HPLHLEDAVTKPELRPCPT 439
   |||:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
   :|:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:

```

```

RESULT 7
US-09-077-955-25
; Sequence 25, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 25
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-25

```

Query Match 21.1%; Score 330; DB 4; Length 456;
 Best Local Similarity 31.3%; Pred. No. 2.8e-22;
 Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPLHGTCQGR-----CLCLPFGSPRCOQSGHGIAESDWHLE 84
 Db 212 DVTSFAGHPCTRASGHFCLNGASCVPREAAVYCLCPGSGPHCEKGL-----VE 261
 QY 85 GSGNDAPQGYAFYFDDGFLAAPPVHRSLSPEVPE-----IELEVR 128
 Db 262 KSAG-----DVTDLAFDGRTFVEYLNNAVTESELANEIPEVKALQSNHPELSLR 309
 QY 129 TSTASGLLLMCGVEVGEAGQKDFISLGLQDGHVFRYOLGSGEARLVSEDPINDGEMHR 188
 Db 310 TEATQGLVLS-----GKATERADYVALAIVDGHQLSYNLGSPVVLRSVTVVNTNRWLR 365
 QY 189 VTALREGRRGSIQVDGEBELVSGRSPGNVAVNAKSVVIGGAPD--VATLTGGRFSSGIT 246
 Db 366 VVAHQREGSLQVGNVAVVTSGLGATQDLDGALMLGGLPELPGVGPALPKAYGTGFV 425
 QY 247 GCVKNLVHSARPGAPPPQPLDLOHRAQAQANTRPCPS 284
 Db 426 GCLRDVVVGR-----HPLHLEDAVTKPELRPCPT 455

RESULT 8
 US-08-644-271-32
 ; Sequence 32, Application US/08644271
 ; Patent No. 5814478
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela, et al.
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
 ; NUMBER OF SEQUENCES: 32
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
 ; STREET: 777 Old Saw Mill Road
 ; CITY: Tarrytown
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10591
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/644,271
 ; FILING DATE: 10-MAY-1996
 ; CLASSIFICATION: 435
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: USSN 60/008,657
 ; FILING DATE: 15-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cobert, Robert J
 ; REGISTRATION NUMBER: 36,108
 ; REFERENCE/DOCKET NUMBER: REG 195A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 914-345-7400
 ; TELEFAX: 914-345-7721
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 32:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 492 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; FRAGMENT TYPE: internal
 ; FEATURE:
 ; NAME/KEY: Human Agrin
 ; LOCATION: 1...492

OTHER INFORMATION:

US-08-644-271-32
 Query Match 21.1%; Score 330; DB 2; Length 492;
 Best Local Similarity 31.3%; Pred. No. 3.1e-22;
 Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPLHGTCQGR-----CLCLPFGSPRCOQSGHGIAESDWHLE 84
 Db 248 DVTSFAGHPCTRASGHFCLNGASCVPREAAVYCLCPGSGPHCEKGL-----VE 297
 QY 85 GSGNDAPQGYAFYFDDGFLAAPPVHRSLSPEVPE-----IELEVR 128
 Db 298 KSAG-----DVTDLAFDGRTFVEYLNNAVTESELANEIPEVKALQSNHPELSLR 345
 QY 129 TSTASGLLLMCGVEVGEAGQKDFISLGLQDGHVFRYOLGSGEARLVSEDPINDGEMHR 188
 Db 346 TEATQGLVLS-----GKATERADYVALAIVDGHQLSYNLGSPVVLRSVTVVNTNRWLR 401
 QY 189 VTALREGRRGSIQVDGEBELVSGRSPGNVAVNAKSVVIGGAPD--VATLTGGRFSSGIT 246
 Db 402 VVAHQREGSLQVGNVAVVTSGLGATQDLDGALMLGGLPELPGVGPALPKAYGTGFV 461
 QY 247 GCVKNLVHSARPGAPPPQPLDLOHRAQAQANTRPCPS 284
 Db 462 GCLRDVVVGR-----HPLHLEDAVTKPELRPCPT 491

RESULT 9
 US-09-077-955-36
 ; Sequence 36, Application US/09077955A
 ; Patent No. 6413740
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela et al., David M.
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
 ; FILE REFERENCE: REG195-B-PCT-US
 ; CURRENT APPLICATION NUMBER: US/09/077,955A
 ; EARLIER FILING DATE: 1998-09-10
 ; EARLIER FILING DATE: 1996-12-13
 ; EARLIER FILING DATE: 08/644,271
 ; EARLIER FILING DATE: 1996-05-10
 ; EARLIER FILING DATE: 60/008,657
 ; EARLIER FILING DATE: 1995-12-15
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO: 36
 ; LENGTH: 492
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; ORGANISM: 09-077-955-36

Query Match 21.1%; Score 330; DB 4; Length 492;
 Best Local Similarity 31.3%; Pred. No. 3.1e-22;
 Matches 87; Conservative 34; Mismatches 99; Indels 58; Gaps 8;

QY 31 DLCEHENPCQLR--EPLHGTCQGR-----CLCLPFGSPRCOQSGHGIAESDWHLE 84
 Db 248 DVTSFAGHPCTRASGHFCLNGASCVPREAAVYCLCPGSGPHCEKGL-----VE 297
 QY 85 GSGNDAPQGYAFYFDDGFLAAPPVHRSLSPEVPE-----IELEVR 128
 Db 298 KSAG-----DVTDLAFDGRTFVEYLNNAVTESELANEIPEVKALQSNHPELSLR 345
 QY 129 TSTASGLLLMCGVEVGEAGQKDFISLGLQDGHVFRYOLGSGEARLVSEDPINDGEMHR 188
 Db 346 TEATQGLVLS-----GKATERADYVALAIVDGHQLSYNLGSPVVLRSVTVVNTNRWLR 401
 QY 189 VTALREGRRGSIQVDGEBELVSGRSPGNVAVNAKSVVIGGAPD--VATLTGGRFSSGIT 246
 Db 402 VVAHQREGSLQVGNVAVVTSGLGATQDLDGALMLGGLPELPGVGPALPKAYGTGFV 461
 QY 247 GCVKNLVHSARPGAPPPQPLDLOHRAQAQANTRPCPS 284

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-31

```

```

Query Match 16.7%; Score 261; DB 4; Length 216;
Best Local Similarity 32.8%; Pred. No. 2.4e-16;
Matches 65; Conservative 27; Mismatches 76; Indels 30; Gaps 4;
QY 105 LAPPGHVFSRSLPEVPEP-----IELEVRTSTASGLLLMQGVGEAGQ 148
Db 30 LAFDGRTFVEYLNATESELANEIPVEKALQSNHFLSLRTEATQGLVLS----GKATE 85
QY 149 GKDFISLGLQDGHVFRYQLGSGEARLIVSEDPINDGWHRVTLALREGRRGSIQVDGELY 208
Db 86 RADIYVALAIVDGHQLQSYNLGSPVWLRSTVFNTRNLKRVVAHREORGSLOVNEAPV 145
QY 209 SGRSPGNVAVNAKGSVVIAGAPD--VATLTGGRFSSGITCVKLVLSHARPAPPPQP 266
Db 146 TGSSPLGATQDLDGALWLGPELPGPALPKAYGTGFCLEDRVVVGR-----HP 197
QY 267 LDLQHRRAQANTREPCS 284
Db 198 LHLLEDVATKPELRPCPT 215

```

```

RESULT 12
US-09-191-647-2
; Sequence 2, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/191,647
; EARLIER FILING DATE: 1998-11-13
; EARLIER APPLICATION NUMBER: 60/065,544
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: 60/081,057
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-191-647-2

```

```

Query Match 16.2%; Score 253.5; DB 3; Length 1525;
Best Local Similarity 24.8%; Pred. No. 1.6e-14;
Matches 92; Conservative 48; Mismatches 100; Indels 131; Gaps 18;
QY 1 CERPCQHGATC-MPAGEBYE--FQCLCRDRGFKDIJCE-----HEENPC----- 40
Db 957 CISNFKKGGTCHLKEGSEDFWCICADGFEGENCEVNVDDCEDNCCENNSTCVDGINNY 1016
QY 41 -----QLREPCCLHGTC-----QSTRCLCLPGFSGP----- 66
Db 1017 TCLCPPEVTGELCEKLDPCAQDLNFCQHDSKLLITPKFKCDCTPGVVEHCDIDFDDC 1076
QY 67 ----RCQSGSGHIAESDWHL-----EGSGG-----NDA----- 91
Db 1077 QDNCKKNGAHTDVAWNGYTCICPEGYSGLFCFEFSPMVLPRTPFCDFNFCQNGAQCIVRI 1136
QY 92 -----FGOYGA-----YFHDGFLAFPGHVSRSLSPEVPT-IELEVRTSTAS 133
Db 1137 NEPICQIPGYQGEKCKLVSNVFNKESYLQIP-----SAKVRPQNTITLQIATDEDS 1190
QY 134 GLLLMQGVGEAGGQKDFISLQDGHVFRYQLGSGEARLV-SEDPINDGWHRVTL 192
Db 1191 GILLYKG-----DKDHIAVELYRGRVRSASVDTGSHPASAIYSVETINDGNFVELL 1242

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-31

```

```

Query Match 20.8%; Score 326; DB 4; Length 256;
Best Local Similarity 32.1%; Pred. No. 3e-22; Indels 56; Gaps 7;
Matches 84; Conservative 32; Mismatches 90; Indels 56; Gaps 7;
QY 45 PCLHGTCQGTR----CLCLPGFSPRCQSGSGHIAESDWHLEGGGNDAPGQYGFYH 100
Db 28 PCLNGASCVREAAVYVCLCPGFGSPHCCKGL-----VEKSGA----- 65
QY 101 DDGFLAPGHVFSRSLPEVPEP-----IELEVRTSTASGLLLMQGVYEVG 144
Db 66 DVDTLAFDGRTFVEYLNATESELANEIPVEKALQSNHFLSLRTEATQGLVLS----G 121
QY 145 EAGQGKDFISLGLQDGHVFRYQLGSGEARLIVSEDPINDGWHRVTLALREGRRGSIQVDG 204
Db 122 KATERADYVALAIVDGHQLQSYNLGSPVWLRSTVFNTRNLKRVVAHREORGSLOVNGN 181
QY 205 EELVSRSPGNVAVNAKGSVVIAGAPD--VATLTGGRFSSGITCVKLVLSHARPAP 262
Db 182 EAPVTGSSPLGATQDLDGALWLGPELPGPALPKAYGTGFCLEDRVVVGR----- 235
QY 263 PPOQLDQHRRAQANTREPCS 284
Db 236 --HPLHLEDAVTKPELRPCPT 255

```

```

RESULT 11
US-09-077-955-31
; Sequence 31, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 31
; LENGTH: 216

```

```

Query Match 20.8%; Score 326; DB 4; Length 256;
Best Local Similarity 32.1%; Pred. No. 3e-22; Indels 56; Gaps 7;
Matches 84; Conservative 32; Mismatches 90; Indels 56; Gaps 7;
QY 45 PCLHGTCQGTR----CLCLPGFSPRCQSGSGHIAESDWHLEGGGNDAPGQYGFYH 100
Db 28 PCLNGASCVREAAVYVCLCPGFGSPHCCKGL-----VEKSGA----- 65
QY 101 DDGFLAPGHVFSRSLPEVPEP-----IELEVRTSTASGLLLMQGVYEVG 144
Db 66 DVDTLAFDGRTFVEYLNATESELANEIPVEKALQSNHFLSLRTEATQGLVLS----G 121
QY 145 EAGQGKDFISLGLQDGHVFRYQLGSGEARLIVSEDPINDGWHRVTLALREGRRGSIQVDG 204
Db 122 KATERADYVALAIVDGHQLQSYNLGSPVWLRSTVFNTRNLKRVVAHREORGSLOVNGN 181
QY 205 EELVSRSPGNVAVNAKGSVVIAGAPD--VATLTGGRFSSGITCVKLVLSHARPAP 262
Db 182 EAPVTGSSPLGATQDLDGALWLGPELPGPALPKAYGTGFCLEDRVVVGR----- 235
QY 263 PPOQLDQHRRAQANTREPCS 284
Db 236 --HPLHLEDAVTKPELRPCPT 255

```

Qy 193 REGRGSIQVDS--EELVSRSGPNNVAVNAKGSVYIGAP---DVATL--TGRFSSGI 245
 Db 1243 ALDQSLSLVDGGNPKIITNLS--KQSTLNFDSPLYVGGMPGKSNVSLRQAPGQNGTSF 1300
 Qy 246 TGCVKNLVLS 256
 Db 1301 HGCIRNLYINS 1311

RESULT 13
 US-09-540-245A-2
 ; Sequence 2, Application US/09540245A
 ; Patent No. 6270984
 ; GENERAL INFORMATION:
 ; APPLICANT: Goodman, Corey
 ; APPLICANT: Kid, Thomas
 ; APPLICANT: Brose, Katja
 ; APPLICANT: Tessier-Lavigne, Marc
 ; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
 ; FILE REFERENCE: B98-031-3
 ; CURRENT APPLICATION NUMBER: US/09/540,245A
 ; CURRENT FILING DATE: 2000-03-31
 ; PRIOR APPLICATION NUMBER: 60/065,544
 ; PRIOR FILING DATE: 1997-11-14
 ; PRIOR APPLICATION NUMBER: 60/081,057
 ; PRIOR FILING DATE: 1998-04-07
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 1525
 ; TYPE: PRT
 ; ORGANISM: human
 US-09-540-245A-2

Query Match 16.2%; Score 253.5; DB 3; Length 1525;
 Best Local Similarity 24.8%; Pred. No. 1.6e-14;
 Matches 92; Conservative 48; Mismatches 100; Indels 131; Gaps 18;

Qy 1 CEROPCOHGATC-MPAGEYE-FOCLCRDGFKGLDCE-----HLENPC----- 40
 Db 957 CISNPKHGHTCHLKEGEDGFWCICADGFEGENCEVNVDDCEDNCCENNSTCVDGINNY 1016

Qy 41 -----QLREPCLLHGTC-----CGTRCLCLPGFSGP----- 66
 Db 1017 TCLCPPEYTGELCEKLDLFCADLNPQHDKILTPKFKDCTPGYVGEHCDDIDFDDC 1076

Qy 67 ----RCQSGSHGIAESDWHL----EGSGG-----NDA----- 91
 Db 1077 QDNCKNGAHTDAVNGYTCICPEGYSGLFCFEFSPMVLPRSPDNFDCQNGAQCIVRI 1136

Qy 92 -----PGQYGA-----YFHDDGFLAFPGHVFGRSLPEVPET-IELEVRTSTAS 133
 Db 1137 NEPICQLPGYQGEKCEKLVSNFINKESYLQIP-----SAKVRPQITWITLQIATDSDS 1190

Qy 134 GLLLMQGVVEGAGCGKDFISLGLQDHLVFRYQLGSGEARLV-SEDPINDGEWHRYTAL 192
 Db 1191 GILLYKG-----DKDHLAVELYRGRVRSYDTGSHSPASAIYSVETINDGNPHIVELL 1242

Qy 193 REGRGSIQVDS--EELVSRSGPNNVAVNAKGSVYIGAP---DVATL--TGRFSSGI 245
 Db 1243 ALDQSLSLVDGGNPKIITNLS--KQSTLNFDSPLYVGGMPGKSNVSLRQAPGQNGTSF 1300

Qy 246 TGCVKNLVLS 256
 Db 1301 HGCIRNLYINS 1311

RESULT 14
 US-09-540-153-2
 ; Sequence 2, Application US/09540153
 ; Patent No. 6270995
 ; GENERAL INFORMATION:
 ; APPLICANT: Goodman, Corey
 ; APPLICANT: Kid, Thomas
 ; APPLICANT: Brose, Katja
 ; APPLICANT: Tessier-Lavigne, Marc
 ; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
 ; FILE REFERENCE: B98-031-3
 ; CURRENT APPLICATION NUMBER: US/09/191,647
 ; CURRENT FILING DATE: 1998-11-13
 ; EARLIER APPLICATION NUMBER: 60/065,544
 ; EARLIER FILING DATE: 1997-11-14
 ; EARLIER APPLICATION NUMBER: 60/081,057
 ; EARLIER FILING DATE: 1998-04-07
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 9

APPLICANT: Goodman, Corey
 APPLICANT: Kid, Thomas
 APPLICANT: Brose, Katja
 APPLICANT: Tessier-Lavigne, Marc
 TITLE OF INVENTION: Modulating Robo: Ligand Interactions
 FILE REFERENCE: B98-031-3
 CURRENT APPLICATION NUMBER: US/09/540,153
 CURRENT FILING DATE: 2000-03-31
 PRIOR APPLICATION NUMBER: 09/191,647
 PRIOR FILING DATE: 1998-11-13
 PRIOR APPLICATION NUMBER: 60/081,057
 PRIOR FILING DATE: 1998-04-07
 NUMBER OF SEQ ID NOS: 14
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 2
 LENGTH: 1525
 TYPE: PRT
 ORGANISM: human
 US-09-540-153-2

Query Match 16.2%; Score 253.5; DB 3; Length 1525;
 Best Local Similarity 24.8%; Pred. No. 1.6e-14;
 Matches 92; Conservative 48; Mismatches 100; Indels 131; Gaps 18;

Qy 1 CEROPCOHGATC-MPAGEYE-FOCLCRDGFKGLDCE-----HLENPC----- 40
 Db 957 CISNPKHGHTCHLKEGEDGFWCICADGFEGENCEVNVDDCEDNCCENNSTCVDGINNY 1016

Qy 41 -----QLREPCLLHGTC-----CGTRCLCLPGFSGP----- 66
 Db 1017 TCLCPPEYTGELCEKLDLFCADLNPQHDKILTPKFKDCTPGYVGEHCDDIDFDDC 1076

Qy 67 ----RCQSGSHGIAESDWHL----EGSGG-----NDA----- 91
 Db 1077 QDNCKNGAHTDAVNGYTCICPEGYSGLFCFEFSPMVLPRSPDNFDCQNGAQCIVRI 1136

Qy 92 -----PGQYGA-----YFHDDGFLAFPGHVFGRSLPEVPET-IELEVRTSTAS 133
 Db 1137 NEPICQLPGYQGEKCEKLVSNFINKESYLQIP-----SAKVRPQITWITLQIATDSDS 1190

Qy 134 GLLLMQGVVEGAGCGKDFISLGLQDHLVFRYQLGSGEARLV-SEDPINDGEWHRYTAL 192
 Db 1191 GILLYKG-----DKDHLAVELYRGRVRSYDTGSHSPASAIYSVETINDGNPHIVELL 1242

Qy 193 REGRGSIQVDS--EELVSRSGPNNVAVNAKGSVYIGAP---DVATL--TGRFSSGI 245
 Db 1243 ALDQSLSLVDGGNPKIITNLS--KQSTLNFDSPLYVGGMPGKSNVSLRQAPGQNGTSF 1300

Qy 246 TGCVKNLVLS 256
 Db 1301 HGCIRNLYINS 1311

RESULT 15
 US-09-191-647-9
 ; Sequence 9, Application US/09191647
 ; Patent No. 6046015
 ; GENERAL INFORMATION:
 ; APPLICANT: Goodman, Corey
 ; APPLICANT: Kid, Thomas
 ; APPLICANT: Brose, Katja
 ; APPLICANT: Tessier-Lavigne, Marc
 ; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
 ; FILE REFERENCE: B98-031-3
 ; CURRENT APPLICATION NUMBER: US/09/191,647
 ; CURRENT FILING DATE: 1998-11-13
 ; EARLIER APPLICATION NUMBER: 60/065,544
 ; EARLIER FILING DATE: 1997-11-14
 ; EARLIER APPLICATION NUMBER: 60/081,057
 ; EARLIER FILING DATE: 1998-04-07
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 9

LENGTH: 735
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-09-191-647-9

Query Match 15.9%; Score 249; DB 3; Length 735;
Best Local Similarity 26.7%; Pred. No. 1.6e-14;
Matches 84; Conservative 48; Mismatches 115; Indels 68; Gaps 17;

```

QY 5 PCHGATCPA-GEYEFCLCRDGFKGLCEHENPCOLREPCLHGGTC---QGTRCLC 59
Db 320 PCENNGKCLIPNGSY--SCMCSPGFTGNCTNIDDCNVE-CQNGSCVDGILLSYDCLC 376
QY 60 LFGFSG-----PRCQ---GSGHGIA---ESDWHL---EGSGGNDAPGQ 94
Db 377 RFGYAGQYCEIIPPMDMEYQKTDACQSQAGCGECVASQNSDFTCKCHEGFGSPCDRQ 436
QY 95 YGAYFHDDG-FLAFPGHVFSRSLPEVPE-TIELEVRTSTASGLLWQGVVEGAGQKDF 152
Db 437 MSVGFKNFGAYLALD-----FLASDGTITWLRITTSKIGILLYYQDD-----HF 480
QY 153 ISLGLQDGHVFRYQIGSGEA-RLVSEDPINDGEHWRVTLREGRRSIQVDGE--ELVS 209
Db 481 VSAELYDGRVKLVYVYIGNPPASHMYSSVKVNDGLPHRISIRTSERKCFLOIDKNFVQIVE 540
QY 210 GRSPGNVAVNAKGSVYIGGAP----DVATLTGGRFSSGITGCVKNLVLSARPGAPPP 264
Db 541 NSGKSDQLITKEMLYIGGLEPIEKSDAKRRFHVKNSLKGCISSITINEV----- 593
QY 265 QFLDLQHRAQAGANT 279
Db 594 -PINLQ-QALENVNT 606

```

Search completed: March 9, 2004, 17:23:33
JOB time : 12.1326 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:11:22 ; Search time 34.5822 Seconds
(without alignments)
2320.373 Million cell updates/sec

Title: US-10-006-011a-9
Perfect score: 1566
Sequence: 1 CERQPCQHGATCMPAGEYEF.....QLDLQHRQAQANTRPCPS 284

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001s.*
- 5: Geneseq2002s.*
- 6: Geneseq2003as.*
- 7: Geneseq2003bs.*
- 8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1566	100.0	4391	6	AAE34390	Aae34390 Human per
2	1559	99.6	4393	4	AAAB31889	Aab31889 Amino aci
3	1543	98.5	4436	4	ABG23265	Abg23265 Novel hum
4	1030	65.8	195	4	AAAB31890	Aab31890 Amino aci
5	357.5	22.8	1931	6	ABUS2400	Abus2400 Human GPC
6	343	21.9	238	4	AAU18145	Aau18145 Novel hum
7	343	21.9	238	4	AAU17011	Aau17011 Human nov
8	343	21.9	238	4	ABBI0450	Abbi0450 Human cdn
9	343	21.9	238	4	AAU19961	Aau19961 Novel hum
10	343	21.9	238	5	ABJ05772	Abj05772 Novel hum
11	343	21.9	238	5	ABP67037	Abp67037 Human pol
12	343	21.9	375	4	AAU07421	Aau07421 Novel hum
13	343	21.9	432	4	AAU18102	Aau18102 Novel hum
14	343	21.9	432	4	AAU16938	Aau16938 Human nov
15	343	21.9	432	4	ABBI0233	Abbi0233 Human cdn
16	343	21.9	432	4	AAU19905	Aau19905 Novel hum
17	343	21.9	432	5	ABJ05729	Abj05729 Novel hum
18	343	21.9	432	5	ABP66820	Abp66820 Human pol
19	343	21.9	463	7	ADB64904	Adb64904 Human pro
20	343	21.9	671	6	ABP58231	Abp58231 Human cel
21	343	21.9	1009	7	ADE28105	Ade28105 Human ntr
22	340.5	21.7	406	5	ABP41801	Abp41801 Human ova
23	338.5	21.6	2053	7	ADC39154	Adc39154 Novel hum
24	338.5	21.6	2143	7	ADC39164	Adc39164 Novel hum
25	334	21.3	416	2	AAAY73993	Aay73993 Human pro

26	334	21.3	1565	7	ADC39156	Adc39156 Novel hum
27	334	21.3	1566	7	ADC39166	Adc39166 Novel hum
28	330	21.1	492	2	AAW26609	Aaw26609 Human agr
29	330	21.1	819	5	ABB72291	Abb72291 Rat prote
30	310	19.8	1544	7	ADC39162	Adc39162 Novel hum
31	303	19.3	1741	5	ABF43859	Abf43859 Human msn
32	292	18.6	1298	4	ABB61948	Abb61948 Drosophil
33	292	18.6	1508	2	AAAY27141	Aay27141 Human sli
34	292	18.6	1508	2	AAW96706	Aaw96706 Protein s
35	292	18.6	1508	2	AAAY04138	Aay04138 Human sli
36	292	18.6	1534	2	AAW46966	Aaw46966 Amino aci
37	292	18.6	1534	2	AAAY27144	Aay27144 Human sli
38	292	18.6	1534	2	AAW96707	Aaw96707 Protein s
39	292	18.6	1534	2	AAAY04139	Aay04139 Human sli
40	292	18.6	1534	7	ADES4473	Ades4473 Human pro
41	290.5	18.6	210	4	AAU19782	Aau19782 Human nov
42	290.5	18.6	210	5	ABP48002	Abp48002 Human pol
43	290.5	18.6	210	7	ADC10964	Adc10964 Human ext
44	281	17.9	4072	4	ABB63614	Abb63614 Drosophil
45	279.5	17.8	152	5	ABP56528	Abp56528 Bm hepara

ALIGNMENTS

RESULT 1
AAE34390
ID AAE34390 standard; protein; 4391 AA.

XX AAE34390;

XX 14-MAY-2003 (first entry)

XX Human perlecan protein.

XX Human; diagnosis; osteoarthritis; rheumatoid arthritis; perlecan.

XX Homo sapiens.

XX WO200295415-A2.

XX 28-NOV-2002.

XX 22-MAY-2002; 2002WO-EPO05612.

XX 23-MAY-2001; 2001GB-00012626.

XX (OSTE-) OSTEOMETER BIO TECH AS.

XX Christgau S, Henriksen DB, Cloos PAC;

XX WPI; 2003-140389/13.

XX An assay for the diagnosis or assessment of the severity of osteoarthritis or rheumatoid arthritis comprising detecting an isomerized or optically inverted protein in a sample.

XX Disclosure; Page 46-67; 106pp; English.

XX The invention relates to an assay for the diagnosis or assessment of the severity of osteoarthritis or rheumatoid arthritis. The assay involves measuring (in a biological sample) the amount or presence of an isomerized or optically inverted protein or one or more isomerized or optically inverted fragments from proteins such as perlecan, biglycan, decorin, fibrillin-1 or procadherin. The assay is useful for the diagnosis or assessment of the severity of osteoarthritis or rheumatoid arthritis. The present sequence is human perlecan protein

XX Sequence 4391 AA;

Query Match 100.0%; Score 1566; DB 6; Length 4391;

Best Local Similarity 100.0%; Pred. No. 8.2e-110;

Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 4393 AA;

Query Match 99.68; Score 1559; DB 4; Length 4393;
 Best Local Similarity 99.38; Pred. No. 2.8e-109;
 Matches 282; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CEROPCHGATCMPAGEYEFQCLCRDGFKGDLCHEHENPCQLREPCFLHGGTCQGRCLCL 60
 Db 4108 CEROPCHGATCMPAGEYEFQCLCRDGFKGDLCHEHENPCQLREPCFLHGGTCQGRCLCL 4167
 QY 61 PFGSPRCQCGSGHGIAESDWHLESGGNDAPQYGYAFHDDGFLAFPPGHVFSRSLPEVP 120
 Db 4168 PFGSPRCQCGSGHGIAESDWHLESGGNDAPQYGYAFHDDGFLAFPPGHVFSRSLPEVP 4227
 QY 121 ETIELEVRTSTASGLLLWQGVVEGAGQKDFISLGLQDGLHVFYQLGSGEARLVSEDP 180
 Db 4228 ETIELEVRTSTASGLLLWQGVVEGAGQKDFISLGLQDGLHVFYQLGSGEARLVSEDP 4287
 QY 181 INDEGHRVHTALREGRGSIQVDGELVSGSPGNVAVNAKGSVYIGGAPDVATLTGGR 240
 Db 4288 INDEGHRVHTALREGRGSIQVDGELVSGSPGNVAVNAKGSVYIGGAPDVATLTGGR 4347
 QY 241 FSSGITCVKLVLSARPGAPPPQPLDLQHRQAQAGANTRPCPS 284
 Db 4348 FSSGITCVKLVLSARPGAPPPQPLDLQHRQAQAGANTRPCPS 4391

RESULT 2
 AAB31889 standard; protein; 4393 AA.
 AAB31889;
 15-MAY-2001 (first entry)
 Amino acid sequence of a human protein.

Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 rheumatoid polyarthritis; lupus erythematosus; gene therapy.

Homo sapiens.
 WO200105422-A2.
 25-JAN-2001.
 17-JUL-2000; 2000WO-FR002057.
 15-JUL-1999; 99FR-00009372.
 (INMR) BIOMERIEUX STELHYS.

Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 WPI; 2001-159475/16.
 Detecting, preventing and treating degenerative, neurological and
 autoimmune diseases, particularly multiple sclerosis, using specified
 polypeptides or related nucleic acid or ligand.

Claim 1; Page 138-152; 209pp; French.
 The present sequence represents a human protein, which is used in the
 method of the invention. The specification describes a method which uses
 at least one polypeptide or polynucleotide sequence belonging to the
 perlecan, precursor of the retinol-binding plasma protein, precursor of
 the ganglioside GM2 activator, calgranulin B or saposin B protein
 families. The method is used for detecting, preventing or treating a
 degenerative, neurological and/or auto-immune disease. The
 polynucleotides and polypeptides are used for diagnosis, prognosis,
 prevention and treatment of multiple sclerosis (in its various forms and
 phases). They may also be useful in cases of e.g. Alzheimer's and
 Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 polyarthritis and lupus erythematosus, including use as vaccines and in
 gene therapy (expression of sense or antisense sequences). They can also

CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 4393 AA;

Query Match 99.68; Score 1559; DB 4; Length 4393;
 Best Local Similarity 99.38; Pred. No. 2.8e-109;
 Matches 282; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CEROPCHGATCMPAGEYEFQCLCRDGFKGDLCHEHENPCQLREPCFLHGGTCQGRCLCL 60
 Db 4110 CEROPCHGATCMPAGEYEFQCLCRDGFKGDLCHEHENPCQLREPCFLHGGTCQGRCLCL 4169
 QY 61 PFGSPRCQCGSGHGIAESDWHLESGGNDAPQYGYAFHDDGFLAFPPGHVFSRSLPEVP 120
 Db 4170 PFGSPRCQCGSGHGIAESDWHLESGGNDAPQYGYAFHDDGFLAFPPGHVFSRSLPEVP 4229
 QY 121 ETIELEVRTSTASGLLLWQGVVEGAGQKDFISLGLQDGLHVFYQLGSGEARLVSEDP 180
 Db 4230 ETIELEVRTSTASGLLLWQGVVEGAGQKDFISLGLQDGLHVFYQLGSGEARLVSEDP 4289
 QY 181 INDEGHRVHTALREGRGSIQVDGELVSGSPGNVAVNAKGSVYIGGAPDVATLTGGR 240
 Db 4290 INDEGHRVHTALREGRGSIQVDGELVSGSPGNVAVNAKGSVYIGGAPDVATLTGGR 4349
 QY 241 FSSGITCVKLVLSARPGAPPPQPLDLQHRQAQAGANTRPCPS 284
 Db 4350 FSSGITCVKLVLSARPGAPPPQPLDLQHRQAQAGANTRPCPS 4393

RESULT 3
 ABG23265 standard; protein; 4436 AA.
 ABG23265;
 18-FEB-2002 (first entry)
 Novel human diagnostic protein #23256.

Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.
 WO200175067-A2.
 11-OCT-2001.
 30-MAR-2001; 2001WO-US008631.
 31-MAR-2000; 2000US-00540217.
 23-AUG-2000; 2000US-00649167.
 (HYSE-) HYSEQ INC.
 Drmanac RT, Liu C, Tang YT;
 WPI; 2001-639362/73.
 N-PSDB; AAS87452.

New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits and to assess
 biodiversity.
 Claim 20; SEQ ID NO 53624; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II)
 sequences. (I) is useful as hybridisation probes, polymerase chain
 reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 and in recombinant production of (II). The polynucleotides are also used
 in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activities. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 4436 AA;

Query Match 98.5%; Score 1543; DB 4; Length 4436;
 Best Local Similarity 96.2%; Pred. No. 4.7e-108;
 Matches 282; Conservative 0; Mismatches 1; Indels 10; Gaps 1;

QY 1 CERQPCQCATMPAGEYEFQCLCRDGPGLDCEHEENPCOLRPECLHGTCQGRCLCL 60
 DB 1 CERQPCQCATMPAGEYEFQCLCRDGPGLDCEHEENPCOLRPECLHGTCQGRCLCL 4202
 QY 61 PFGSPRCQOQSGHGIASDWHLEGGNDAPGQYGFYHDDGFLAFPGHVFGRSLPEVP 120
 DB 4203 PFGSPRCQOQSGHGIASDWHLEGGNDAPGQYGFYHDDGFLAFPGHVFGRSLPEVP 4262
 QY 121 ETIELEVRTSTASGLLWQVEVGBAGQKDFISLGLQDGHVLF-----RYQLGS 170
 DB 4263 ETIELEVRTSTASGLLWQVEVGBAGQKDFISLGLQDGHVLF-----RYQLGS 4322
 QY 171 GEARLVSEDPINDGEHWRVTALREGRGSIQVDSBELVSGRSPGNVAVNAKGSVIGGA 230
 DB 4323 GEARLVSEDPINDGEHWRVTALREGRGSIQVDSBELVSGRSPGNVAVNAKGSVIGGA 4382
 QY 231 PDVAVLTGRRSSGITGCVKLVLSHSPGAPPDPPDLDLQHRAQAGANTRPCP 283
 DB 4383 PDVAVLTGRRSSGITGCVKLVLSHSPGAPPDPPDLDLQHRAQAGANTRPCP 4435

RESULT 4
 AAB31890
 ID AAB31890 standard; protein; 195 AA.
 XX AAB31890;

XX DT 15-MAY-2001 (first entry)
 XX DE Amino acid sequence of the C-terminal of the human perlecan protein.
 XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 XX KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 XX KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 XX KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 XX KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX OS Homo sapiens.
 XX XX WC200105422-A2.
 XX PN 25-JAN-2001.
 XX PD 17-JUL-2000; 2000MO-FR002057.
 XX XX 15-JUL-1999; 99FR-00009372.
 XX PR (INNR) BIOMERIEUX STELHYS.
 XX PA Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Ferron H;
 XX PI

XX WPI; 2001-159475/16.
 DR N-PSDB; AAF54728.
 XX PT Detecting, preventing and treating degenerative, neurological and
 XX auto-immune diseases, particularly multiple sclerosis, using specified
 XX polypeptides or related nucleic acid or ligand.
 XX Claim 1; Page 152-153; 209pp; French.

The present sequence represents a human polypeptide, which is used in the
 method of the invention. The specification describes a method which uses
 at least one polypeptide or polynucleotide sequence belonging to the
 perlecan, precursor of the retinol-binding plasma protein, precursor of
 the ganglioside GM2 activator, calgranulin B or saposin B protein
 families. The method is used for detecting, preventing or treating a
 degenerative, neurological and/or auto-immune disease. The
 polynucleotides and polypeptides are used for diagnosis, prognosis,
 prevention and treatment of multiple sclerosis (in its various forms and
 phases). They may also be used in cases of e.g. Alzheimer's and
 Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 polyarthritis and lupus erythematosus, including use as vaccines and in
 gene therapy (expression of sense or antisense sequences). They can also
 be used to assess efficacy of potential therapeutic agents, particularly
 compounds that reduce or inhibit toxicity towards glial cells

XX SQ Sequence 195 AA;
 Query Match 65.8%; Score 1030; DB 4; Length 195;
 Best Local Similarity 100.0%; Pred. No. 1.5e-70;
 Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 DAPQGYGFHDDGFLAFPGHVFGRSLPEVPTIELEVRTSTASGLLWQVEVGBAGQ 149
 DB 1 DAPQGYGFHDDGFLAFPGHVFGRSLPEVPTIELEVRTSTASGLLWQVEVGBAGQ 60
 QY 150 KDFISLGLQDGHVLFYQLGSGEARLVSEDPINDGEHWRVTALREGRGSIQVDSBELV 209
 DB 51 KDFISLGLQDGHVLFYQLGSGEARLVSEDPINDGEHWRVTALREGRGSIQVDSBELV 120
 QY 210 GRSPGNVAVNAKGSVIGGAPPDPPDLDLQHRAQAGANTRPCP 283
 DB 121 GRSPGNVAVNAKGSVIGGAPPDPPDPPDLDLQHRAQAGANTRPCP 195

RESULT 5
 ABU52400
 ID ABU52400 standard; protein; 1931 AA.
 XX ABU52400;
 XX DT 03-MAR-2003 (first entry)
 XX DE Human GPCR related protein NOV40a.
 XX KW Human; NOVX; G-protein coupled receptor; GPCR; cancer; cytosstatic.
 XX OS Homo sapiens.
 XX PN WO200279398-A2.
 XX PD 10-OCT-2002.
 XX XX 08-MAR-2002; 2002WO-US007355.
 XX PR 08-MAR-2001; 2001US-0274194P.
 XX PR 08-MAR-2001; 2001US-0274281P.
 XX PR 08-MAR-2001; 2001US-0274322P.
 XX PR 09-MAR-2001; 2001US-0274849P.

Matches	91; Conservative	29; Mismatches	110; Indels	27; Gaps	8;
QY	1	CEBOPCOHGATCMEAGEVEFQCLCRDGFKGDLCHEHENPCOLREPCILHGTCQ----	GT	55	
Db	1439	CLPNECHGAFQNLKAGRFHCQCPGRVGTCADEKSPCQ-FNPCHGAAPCRVLP	EGGA	1497	
QY	56	RCLCLPFGSPRCQGGSHGIAESDWELEGGNDAPQYGAFFHDDGFLAPFG-HVFSR	114		
Db	1498	QCECPGREGTFCQTAS-----GQDSSGPFADFNFGFSHLELRLHTFFAR	1542		
QY	115	SUPEVPEITIELEVRTSTASGILLMQGVEVGEAGCGKDFISLGLQDGHVFRYQLGSEAR	174		
Db	1543	DLGE-KMALEVVVFLARGPSGLLYNGKTD--GKG-DFVSLALDRRLRLEFRYDLGKAAV	1598		
QY	175	LVSEDPINDGEMHWYALREGRGSIQVDGEELVSGSPGNVAVNAKGSVYIGGAPDVA	234		
Db	1599	IKSREFVTLGAWTRVSLERNGKALRVGDFRVLGSPVPHVTLKKEFLYVGGAPDFS	1658		
QY	235	TIT-GGFSSGITCVK 250			
Db	1659	KLARAAVSSGFGAIIQ 1675			
RESULT 6					
AAU18145	AAU18145 standard; protein; 238 AA.				
XX	AC	AAU18145;			
XX	DT	21-NOV-2001 (first entry)			
XX	DE	Novel human uterine motility-association polypeptide #52.			
XX	KW	Human; uterine motility-association disorder; uterus; pregnancy; labour; menstrual cycle; gene therapy.			
XX	OS	Homo sapiens.			
XX	PN	WO2001155201-A1.			
XX	XX	02-AUG-2001.			
XX	PF	17-JAN-2001; 2001WO-US001317.			
XX	PR	31-JAN-2000; 2000US-0179065P.			
XX	PR	04-FEB-2000; 2000US-0180628P.			
XX	PR	24-FEB-2000; 2000US-0184664P.			
XX	PR	02-MAR-2000; 2000US-0186350P.			
XX	PR	16-MAR-2000; 2000US-0189874P.			
XX	PR	17-MAR-2000; 2000US-0190076P.			
XX	PR	18-APR-2000; 2000US-0196123P.			
XX	PR	19-MAY-2000; 2000US-0205515P.			
XX	PR	07-JUN-2000; 2000US-0209467P.			
XX	PR	28-JUN-2000; 2000US-0214886P.			
XX	PR	30-JUN-2000; 2000US-0215135P.			
XX	PR	07-JUL-2000; 2000US-0216647P.			
XX	PR	11-JUL-2000; 2000US-0216880P.			
XX	PR	11-JUL-2000; 2000US-0217487P.			
XX	PR	14-JUL-2000; 2000US-0217496P.			
XX	PR	26-JUL-2000; 2000US-0220963P.			
XX	PR	14-AUG-2000; 2000US-0224518P.			
XX	PR	14-AUG-2000; 2000US-0224519P.			
XX	PR	14-AUG-2000; 2000US-0225213P.			
XX	PR	14-AUG-2000; 2000US-0225214P.			
XX	PR	14-AUG-2000; 2000US-0225256P.			
XX	PR	14-AUG-2000; 2000US-0225267P.			
XX	PR	14-AUG-2000; 2000US-0225268P.			
XX	PR	14-AUG-2000; 2000US-0225270P.			
XX	PR	14-AUG-2000; 2000US-0225447P.			
XX	PR	14-AUG-2000; 2000US-0225757P.			
XX	PR	14-AUG-2000; 2000US-0225758P.			

Claim 1; Page 223; 413pp; English.

The present invention relates to the isolation of novel human polypeptides referred to as NOVX (NOVI-NOV44), variants of these proteins, and the polynucleotide sequences encoding them. The NOVX proteins of the invention are G-protein coupled receptor (GPCR) related proteins. The sequences of the invention are useful in the manufacture of a medicament for treating a syndrome related to a human disease associated with the polypeptides e.g. cancer. ABUS2311-ABUS2408 represent the human NOVX proteins of the invention

Sequence 1931 AA;

Query Match 22.8%; Score 357.5; DB 6; Length 1931;

Best Local Similarity 35.4%; Pred. No. 3.1e-18;

```

PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229511P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 02-OCT-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239938P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 08-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246603P.
PR 08-NOV-2000; 2000US-0246610P.

```

```

PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250150P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251956P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-02559678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-488777/53.
XX N-P9DB; AAS28987.
XX
XX New nucleic acid molecules encoding 49 human secreted proteins for
XX diagnosing, preventing, treating or ameliorating medical conditions and
XX used as food additives or preservatives.
XX
XX Claim 11; SEQ ID NO 121; 524pp; English.
XX
XX The present invention relates to the isolation of novel human uterine
XX motility-association polypeptides, and cDNA (AAS28936-AAS28994) and
XX genomic sequences encoding for these polypeptides. The sequences of the
XX invention are useful in the diagnosis, treatment, prevention and/or
XX prognosis of diseases associated with uterine motility such as pregnancy
XX and labour, and menstrual disorders. The polynucleotide sequences of the
XX invention are also useful in gene therapy. AAU18094-AAU18152 represent
XX novel human uterine motility-association polypeptides. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 238 AA;

```

```

Query Match 21.9%; Score 343; DB 4; Length 238;
Best Local Similarity 33.5%; Pred. No. 3.6e-18;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

OY 34 EHENPCOLRPCLHGTC----QGRCLCLPFGSFCRCQQSGHGSIABSDMHLGSSGN 89
Db : : : : : : : : : : : : : : : : : : : : : : : : :
11 ENAAHPG-VRAPCAHGGSFRCRPEKGYDCCPLGFEGLHCQKAIIEAI-----56
OY 90 DAP----GOYGFVHDGFLAFEGHVFVRSIPVPTIELVRTSTASGLLLWGVVEVCE 145
Db : : : : : : : : : : : : : : : : : : : : : : : : :
57 EIQFIGRSYLTNDNPILKRVG---SRS-----NVFMFKTKTDGLLLWRG--DSP 105

```

Oy 146 ACQKDFISLGHODHLVFRYOLGSGEARLVSEDPINDEWHRVTAERGRRSIQVDGE 205
 Db 106 MRNSDFISLGRDGLVSYNLSGVSIMVNGSFNDRHRVAVRVDGSGKITVDDY 165
 Oy 206 ELVSGRSPGNVAVNAKGVYIGGAPDVATLTGGRFSSGITCVKNLVLHSARFGAPPQP 265
 Db 166 GARTKSPGMMRQLNTGALYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DY 218
 Oy 266 PLDLQHQRAQAQANTREC 282
 Db 219 HISLVEDAVDCKMINTC 235

RESULT 7
 AAU17011
 ID AAU17011 standard; protein; 238 AA.

XX AAU17011;

DT 07-NOV-2001 (first entry)

XX Human novel secreted protein, SEQ ID 252.

Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
 Cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
 antibacterial; virucide; fungicide; ophthalmological; vulnery;
 secreted protein; rheumatoid arthritis; hyperproliferative disorder;
 cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 cerebral ischaemia; angiogenesis; nervous system disorder;
 Alzheimer's disease; infection; ocular disorder; corneal infection;
 wound healing; epithelial cell proliferation; skin ageing; food additive;
 preservative; antiproliferative.

XX Homo sapiens.

XX WC200155441-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WC-US001320.

XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184654P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216847P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226888P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234224P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246603P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.

PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0255719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0255678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 FI WPI; 2001-476222/51.
 XX N-PSDB; AAS26916.
 DR
 DR Novel polypeptides and polynucleotides useful as diagnostic reagents to
 PT diagnose diseases or disorders associated with aberrant expression or
 PT activity of polypeptides, for treating blood clotting disorder,
 PT hemophilia.
 XX
 XX Claim 11; SEQ ID NO 252; 601pp; English.
 PS
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 CC and many other disorders listed in the specification. The polypeptides
 CC can also be used to aid wound healing and epithelial cell proliferation,
 CC to prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The

Query Match 21.9%; Score 343; DB 4; Length 238;
 Best Local Similarity 33.5%; Pred. No. 3.6e-18;

Matches	86;	Conservative	34;	Mismatches	97;	Indels	40;	Gaps	8;
QY	34	EHEENPQRLREFCLHGQPC	----	QOTRCLCLPFGSPRCQQGSGHGIAESDWHLESGGN	89				
Db	11	ENAAHPC-VRAPCAHGSSCRKEGYDCDCPLGFEGLHCKKAILEA	----		56				
QY	90	DAP----	GOYGAYFHDDGFLAPFGHVFSRSLPEVPEETIELEVRTSTASGLLLWQGVGE	145					
Db	57	EIPQFIGRSYLYDNPDLKRVSG	----	SRS-----NVFMRFKYAKDGLLLWRG	105				
QY	146	AGQKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEHWRVYALRREGRRGSIQVDGE	205						
Db	106	MRPNSDFISLGLRDCALVFSYNLGSGVASIMVNGSFNDRGRWHRVYKAVRDCGSGKITVDDY	165						
QY	206	ELVSGRSRQPNVAVNAKGSVVIIGAPDVAVLTGGRFSSGITGVKLVLSARPGAPPQ	265						
Db	166	GARTKSPQMRQLNINALVYVGMKEALHTNQYMRGLVGCISHFTLST	----	----	218				
QY	266	PLDLQHRAQAGANTRPC	282						
Db	219	HISLVEDAVDGNKNTC	235						
RESULT 8									
ID	ABB10450	standard; protein; 238 AA.							
XX	AC	ABB10450;							
XX	DT	10-JAN-2002 (first entry)							
XX	DE	Human CDNA SEQ ID NO: 758.							
XX	XX	Human; gene therapy; neural disorder; immune system disorder;							
XX	KW	muscular disorder; reproductive disorder; gastrointestinal disorder;							
XX	KW	pulmonary disorder; cardiovascular disorder; renal disorder;							
XX	OS	proliferative disorder; inflammation.							
XX	PN	Homo sapiens.							
XX	XX	WO200154474-A2.							
XX	PD	02-AUG-2001.							
XX	PF	17-JAN-2001; 2001WO-US001349.							
XX	PR	31-JAN-2000; 2000US-0179065P.							
XX	PR	04-FEB-2000; 2000US-0180628P.							
XX	PR	24-FEB-2000; 2000US-0184664P.							
XX	PR	02-MAR-2000; 2000US-0186350P.							
XX	PR	16-MAR-2000; 2000US-0189874P.							
XX	PR	17-MAR-2000; 2000US-0190076P.							
XX	PR	19-APR-2000; 2000US-0198123P.							
XX	PR	19-MAY-2000; 2000US-0205515P.							
XX	PR	07-JUN-2000; 2000US-0209467P.							
XX	PR	28-JUN-2000; 2000US-0214886P.							
XX	PR	30-JUN-2000; 2000US-0215135P.							
XX	PR	07-JUL-2000; 2000US-0216647P.							
XX	PR	07-JUL-2000; 2000US-0216880P.							
XX	PR	11-JUL-2000; 2000US-0217487P.							
XX	PR	11-JUL-2000; 2000US-0217496P.							
XX	PR	14-JUL-2000; 2000US-0218290P.							
XX	PR	26-JUL-2000; 2000US-0220963P.							
XX	PR	26-JUL-2000; 2000US-0220964P.							
XX	PR	14-AUG-2000; 2000US-0224518P.							
XX	PR	14-AUG-2000; 2000US-0224519P.							
XX	PR	14-AUG-2000; 2000US-0225213P.							
XX	PR	14-AUG-2000; 2000US-0225214P.							
XX	PR	14-AUG-2000; 2000US-0225266P.							
XX	PR	14-AUG-2000; 2000US-0225267P.							
XX	PR	14-AUG-2000; 2000US-0225268P.							
XX	PR	14-AUG-2000; 2000US-0225270P.							
XX	PR	14-AUG-2000; 2000US-0225447P.							

```

PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228928P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231141P.
PR 08-SEP-2000; 2000US-0231141P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232196P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 26-SEP-2000; 2000US-0234998P.
PR 27-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236357P.
PR 29-SEP-2000; 2000US-0236358P.
PR 29-SEP-2000; 2000US-0236359P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.

104-NOV-2000; 2000US-02456609P.
104-NOV-2000; 2000US-02456610P.
104-NOV-2000; 2000US-02456611P.
104-NOV-2000; 2000US-02456612P.
104-NOV-2000; 2000US-02456613P.
104-NOV-2000; 2000US-0245207P.
104-NOV-2000; 2000US-0245208P.
104-NOV-2000; 2000US-0245209P.
104-NOV-2000; 2000US-0249210P.
104-NOV-2000; 2000US-0249211P.
104-NOV-2000; 2000US-0249212P.
104-NOV-2000; 2000US-0249213P.
104-NOV-2000; 2000US-0249214P.
104-NOV-2000; 2000US-0249215P.
104-NOV-2000; 2000US-0249216P.
104-NOV-2000; 2000US-0249217P.
104-NOV-2000; 2000US-0249218P.
104-NOV-2000; 2000US-0249244P.
104-NOV-2000; 2000US-0249245P.
104-NOV-2000; 2000US-0249246P.
104-NOV-2000; 2000US-0249265P.
104-NOV-2000; 2000US-0249297P.
104-NOV-2000; 2000US-0249299P.
01-DEC-2000; 2000US-0250160P.
02-DEC-2000; 2000US-0250391P.
03-DEC-2000; 2000US-0251030P.
05-DEC-2000; 2000US-0251988P.
05-DEC-2000; 2000US-0256719P.
06-DEC-2000; 2000US-0251479P.
08-DEC-2000; 2000US-0251856P.
08-DEC-2000; 2000US-0251869P.
08-DEC-2000; 2000US-0251989P.
08-DEC-2000; 2000US-0251990P.
11-DEC-2000; 2000US-0254097P.
05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;
WPI; 2001-476161/51.
N-PSDB; ABA06672.

Isolated nucleic acid molecule encoding an inflammation-associated
polypeptide is used in preventing, treating or ameliorating a medical
condition.

Claim 11; SEQ ID NO 758; 859pp + Sequence Listing, English.

The present invention provides human cDNAs, proteins and related genomic
DNAs. These can be used in the treatment of neural, immune system,
muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
renal and proliferative disorders and inflammation. The present sequence
is a protein of the invention

Sequence 238 AA;
Query Match 21.9%; Score 143; DB 4; Length 238;
Best Local Similarity 33.5%; Pred No 3; ee-18;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
Qy 34 EHEENPCQLREPCFLHGTC---QGTCLCLLPFGSPRCQSGHGIAESDWHLESGGN 89
Db 11 ENAHPC-VRAPCAHGGSCRRKEGYPDCCPLGFEGLHCKAIEAI----- 56
Qy 90 DAP---GOYGAYFHDGFLAFPGHFVSRSLPEVPEPIELETSTASGLLWQVEVGE 145
Db 57 EIPQFIGRSYLYDNPDLKVEVS---SR-----NVFMRFKTKDGLLRG--DSP 105
Qy 146 ACGGKDFISLGLDCHLVFRYQLQSGEARLVSEDPINDEWHHVTALREGRRSIVDGE 205
Db 106 MRPNSDFISLGLRDLGVYVFNLGSGVASIMVNGSFNDGWRHRVKAIRDQCQSGKITVDDY 165

```

QY	206	ELVSGRGGPNNVAVNAKGSVYIGGAPDVAVLTGGRFSSGITCVKKNLVHHSARGPAPPPQ 265	
Db	166	GARTKSPGNNRQLNINALYVGGMKRIALHTNRQYMRGLVGCISHFTLST-----DY 218	
QY	266	FLDLOHRAQAAGANTRPC 282	
Dd	219	HISLVEDAVDQKNINYC 235	
RESULT 9			
AAU19961			
ID	AAU19961	standard; protein; 238 AA.	
AC	AAU19961;		
DT	06-DEC-2001	(first entry)	
XX	DE	Novel human calcium-binding protein #70.	
XX	KW	Human; calcium-binding protein; calcium flux; neurological disease;	
KW	KW	immune dysfunction; digestive disorder; neoplastic disease;	
KW	KW	blood disorder; infectious disease; Gene therapy; immunosuppressive;	
KW	KW	antiarthritic; cytostatic; vasotropic; antibacterial; nootropic;	
XX	XX	virucide.	
OS	OS	Homo sapiens.	
XX	FN	WO200155304-A2.	
XX	PD	02-AUG-2001.	
XX	PF	17-JAN-2001; 2001WO-US0001302.	
XX	PR	31-JAN-2000; 2000US-0179065P.	
PR	PR	04-FEB-2000; 2000US-0180628P.	
PR	PR	24-FEB-2000; 2000US-0184664P.	
PR	PR	02-MAR-2000; 2000US-0186350P.	
PR	PR	16-MAR-2000; 2000US-0189874P.	
PR	PR	17-MAR-2000; 2000US-0190076P.	
PR	PR	18-APR-2000; 2000US-0198123P.	
PR	PR	19-MAY-2000; 2000US-0205515P.	
PR	PR	07-JUN-2000; 2000US-0209467P.	
PR	PR	28-JUN-2000; 2000US-0214886P.	
PR	PR	30-JUN-2000; 2000US-0215135P.	
PR	PR	07-JUL-2000; 2000US-0216647P.	
PR	PR	11-JUL-2000; 2000US-0216880P.	
PR	PR	11-JUL-2000; 2000US-0217487P.	
PR	PR	14-JUL-2000; 2000US-0217496P.	
PR	PR	14-JUL-2000; 2000US-0218290P.	
PR	PR	26-JUL-2000; 2000US-0220963P.	
PR	PR	26-JUL-2000; 2000US-0220964P.	
PR	PR	14-AUG-2000; 2000US-0224518P.	
PR	PR	14-AUG-2000; 2000US-0224519P.	
PR	PR	14-AUG-2000; 2000US-0225213P.	
PR	PR	14-AUG-2000; 2000US-0225214P.	
PR	PR	14-AUG-2000; 2000US-0225266P.	
PR	PR	14-AUG-2000; 2000US-0225267P.	
PR	PR	14-AUG-2000; 2000US-0225268P.	
PR	PR	14-AUG-2000; 2000US-0225270P.	
PR	PR	14-AUG-2000; 2000US-0225271P.	
PR	PR	14-AUG-2000; 2000US-0225272P.	
PR	PR	14-AUG-2000; 2000US-0225273P.	
PR	PR	14-AUG-2000; 2000US-0225274P.	
PR	PR	14-AUG-2000; 2000US-0225275P.	
PR	PR	14-AUG-2000; 2000US-0225276P.	
PR	PR	14-AUG-2000; 2000US-0225277P.	
PR	PR	14-AUG-2000; 2000US-0226279P.	
PR	PR	22-AUG-2000; 2000US-0226681P.	
PR	PR	22-AUG-2000; 2000US-0226682P.	
PR	PR	22-AUG-2000; 2000US-0227182P.	
PR	PR	23-AUG-2000; 2000US-0227009P.	
PR	PR	30-AUG-2000; 2000US-0228242P.	
PR	PR	01-SEP-2000; 2000US-0229247P.	
PR	PR	01-SEP-2000; 2000US-0229343P.	
PR	PR	01-SEP-2000; 2000US-0229344P.	
PR	PR	01-SEP-2000; 2000US-0229345P.	
PR	PR	05-SEP-2000; 2000US-0229509P.	
PR	PR	05-SEP-2000; 2000US-0229513P.	
PR	PR	06-SEP-2000; 2000US-0230437P.	
PR	PR	06-SEP-2000; 2000US-0230438P.	
PR	PR	08-SEP-2000; 2000US-0231242P.	
PR	PR	08-SEP-2000; 2000US-0231243P.	
PR	PR	08-SEP-2000; 2000US-0231244P.	
PR	PR	08-SEP-2000; 2000US-0231413P.	
PR	PR	08-SEP-2000; 2000US-0231414P.	
PR	PR	08-SEP-2000; 2000US-0232080P.	
PR	PR	12-SEP-2000; 2000US-0232081P.	
PR	PR	14-SEP-2000; 2000US-0232397P.	
PR	PR	14-SEP-2000; 2000US-0232398P.	
PR	PR	14-SEP-2000; 2000US-0232399P.	
PR	PR	14-SEP-2000; 2000US-0232400P.	
PR	PR	14-SEP-2000; 2000US-0232401P.	
PR	PR	14-SEP-2000; 2000US-0233063P.	
PR	PR	14-SEP-2000; 2000US-0233064P.	
PR	PR	21-SEP-2000; 2000US-0233065P.	
PR	PR	21-SEP-2000; 2000US-0234223P.	
PR	PR	21-SEP-2000; 2000US-0234274P.	
PR	PR	25-SEP-2000; 2000US-0234997P.	
PR	PR	25-SEP-2000; 2000US-0234998P.	
PR	PR	26-SEP-2000; 2000US-0235484P.	
PR	PR	27-SEP-2000; 2000US-0235834P.	
PR	PR	27-SEP-2000; 2000US-0235836P.	
PR	PR	29-SEP-2000; 2000US-0236327P.	
PR	PR	29-SEP-2000; 2000US-0236367P.	
PR	PR	29-SEP-2000; 2000US-0236368P.	
PR	PR	29-SEP-2000; 2000US-0236369P.	
PR	PR	29-SEP-2000; 2000US-0236370P.	
PR	PR	02-OCT-2000; 2000US-0236802P.	
PR	PR	02-OCT-2000; 2000US-0237037P.	
PR	PR	02-OCT-2000; 2000US-0237038P.	
PR	PR	02-OCT-2000; 2000US-0237039P.	
PR	PR	02-OCT-2000; 2000US-0237040P.	
PR	PR	13-OCT-2000; 2000US-0239935P.	
PR	PR	13-OCT-2000; 2000US-0239937P.	
PR	PR	20-OCT-2000; 2000US-0240960P.	
PR	PR	20-OCT-2000; 2000US-0241212P.	
PR	PR	20-OCT-2000; 2000US-0241785P.	
PR	PR	20-OCT-2000; 2000US-0241786P.	
PR	PR	20-OCT-2000; 2000US-0241787P.	
PR	PR	20-OCT-2000; 2000US-0241808P.	
PR	PR	20-OCT-2000; 2000US-0241809P.	
PR	PR	01-NOV-2000; 2000US-0241826P.	
PR	PR	08-NOV-2000; 2000US-0244617P.	
PR	PR	08-NOV-2000; 2000US-0244617P.	
PR	PR	08-NOV-2000; 2000US-0246474P.	
PR	PR	08-NOV-2000; 2000US-0246475P.	
PR	PR	08-NOV-2000; 2000US-0246476P.	
PR	PR	08-NOV-2000; 2000US-0246477P.	
PR	PR	08-NOV-2000; 2000US-0246478P.	
PR	PR	08-NOV-2000; 2000US-0246523P.	
PR	PR	08-NOV-2000; 2000US-0246524P.	
PR	PR	08-NOV-2000; 2000US-0246525P.	
PR	PR	08-NOV-2000; 2000US-0246526P.	
PR	PR	08-NOV-2000; 2000US-0246527P.	
PR	PR	08-NOV-2000; 2000US-0246528P.	
PR	PR	08-NOV-2000; 2000US-0246532P.	
PR	PR	08-NOV-2000; 2000US-0246609P.	
PR	PR	08-NOV-2000; 2000US-0246610P.	
PR	PR	08-NOV-2000; 2000US-0246611P.	
PR	PR	08-NOV-2000; 2000US-0246613P.	
PR	PR	17-NOV-2000; 2000US-0249207P.	
PR	PR	17-NOV-2000; 2000US-0249208P.	
PR	PR	17-NOV-2000; 2000US-0249209P.	
PR	PR	17-NOV-2000; 2000US-0249210P.	
PR	PR	17-NOV-2000; 2000US-0249211P.	
PR	PR	17-NOV-2000; 2000US-0249212P.	
PR	PR	17-NOV-2000; 2000US-0249213P.	
PR	PR	17-NOV-2000; 2000US-0249214P.	

17-NOV-2000; 2000US-0249215P.
 17-NOV-2000; 2000US-0249216P.
 17-NOV-2000; 2000US-0249217P.
 17-NOV-2000; 2000US-0249218P.
 17-NOV-2000; 2000US-0249244P.
 17-NOV-2000; 2000US-0249245P.
 17-NOV-2000; 2000US-0249264P.
 17-NOV-2000; 2000US-0249265P.
 17-NOV-2000; 2000US-0249297P.
 17-NOV-2000; 2000US-0249299P.
 17-NOV-2000; 2000US-0249300P.
 01-DEC-2000; 2000US-0250391P.
 05-DEC-2000; 2000US-0251030P.
 05-DEC-2000; 2000US-0251988P.
 06-DEC-2000; 2000US-0256719P.
 08-DEC-2000; 2000US-0251479P.
 08-DEC-2000; 2000US-0251856P.
 08-DEC-2000; 2000US-0251868P.
 08-DEC-2000; 2000US-0251989P.
 08-DEC-2000; 2000US-0251989P.
 11-DEC-2000; 2000US-0254097P.
 05-JAN-2001; 2001US-0259678P.
 XX
 FA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Parash SC, Ruben SM;
 XX
 DR WPI; 2001-465568/50.
 DR N-PSDB; AAS31646.
 XX
 PT Isolated nucleic acid molecule encoding a calcium-binding protein is used
 FT in preventing, treating or ameliorating a medical condition.
 XX
 PF Claim 11; SEQ ID NO 158; 542pp; English.
 XX
 CC The present invention relates to the isolation of novel human calcium-
 CC binding proteins, and cDNA (AAS31577-AAS31654) and genomic sequences
 CC encoding for these proteins. The sequences of the invention are useful in
 CC the diagnosis, prevention and/or prognosis of diseases associated with
 CC aberrant calcium flux. Such disorders include neurological diseases (e.g.
 CC amyotrophic lateral sclerosis, ALS), immune dysfunction (e.g. severe
 CC combined immunodeficiency, SCID), digestive disorders (e.g. irritable
 CC bowel syndrome, IBS), neoplastic disease (e.g. cancer), blood disorders
 CC (e.g. haemophilia), and/or infectious disease (e.g. acquired
 CC immunodeficiency syndrome, AIDS). The novel calcium-binding proteins are
 CC also useful as screening tools to identify antagonists and/or agonists
 CC that may enhance or inhibit activities mediated by calcium-binding
 CC proteins. The polypeptides of the invention are also useful in gene
 CC therapy. AAU1982-AAU1969 represent the novel human calcium-binding
 CC proteins. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 238 AA;

Query Match 21.98; Score 343; DB 4; Length 238;
 Best Local Similarity 33.54; Pred. No. 3.6e-16;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

34 EHENPQLRPLRHGTC---QGRCLCLPFGSFCQSGHGIAEFDWHLRSGGN 89
 11 ENAAHPC-VRAPCAHGSCSRPRKEGYDCDPLGFEGLHCQKAIIEA|----- 56

90 DAP---GQYGAYPHDGLAFPHGTVFVSRSLPEVPTIELEVRTSPASGLLWQGVGE 145
 57 EIPQIGRSYTYDNPDLIKRVSG---SRB-----NVFMRFKTKADGILLWRG--DSP 105

146 AGQCKDFISLQDGHVLFVYQLGSGEARLVSDPINDGEMHWKVTALREGRGSIQVDGE 205
 106 MRPNDFISLCLRDGALVFNGLGSGVASIMWNGSFNDGRWHRVRAVRDQSGKIVDDY 165

QY 206 ELVSGRSPGNVAVNAKGSVYIGGAPVATLGTGRFSSGITCVKMLVLHSARFPAPPQ 265
 DB 166 GARTKSGFQWRLNGLNGALYVCGMKELALHTNRQVNRGLVGCISHFTLST-----DY 218
 QY 266 PLDLQHRAOAGANTRPC 282
 DB 219 HISLVEDAVDGNKINTC 235

RESULT 10
 ABJ05772
 ID ABJ05772 standard; protein; 238 AA.
 AC ABJ05772;
 XX
 DT 14-NOV-2002 (first entry)
 XX
 DE Novel human protein SEQ ID No 121.
 XX
 KW Immunostimulant; antiarthritic; antiarthritic; neuroprotective;
 KW anti-allergic; antidiabetic; antiasthmatic; anti-inflammatory; nootropic;
 KW immunosuppressive; anticoagulant; thrombolytic; antithrombotic;
 KW cytostatic; nephrotropic; antiparkinsonian; gynecological; virucide;
 KW antibacterial; antiarrhythmic; fungicide; HCFAT05; HWAAR95; HTNBM01;
 KW immunodeficiency; autoimmune disorder; allergic reaction; cardiovascular;
 KW inflammatory condition; graft-versus-host disease; reproductive system;
 KW blood-related disorder; hyperproliferative; endocrine; neurological;
 KW respiratory; renal; infectious disease; gastrointestinal; gene therapy;
 KW neuronal growth; neuronal disorder; neuro-degenerative condition;
 KW keratinocyte growth; human.
 XX
 OS Homo sapiens.
 XX
 XX US2002086330-A1.
 XX
 XX 04-JUL-2002.
 XX
 XX 17-JAN-2001; 2001US-00764893.
 XX
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 29-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0218880P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 11-JUL-2000; 2000US-0218290P.
 PR 14-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0228287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234233P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.

PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0245299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.

XX (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.

XX Rosen CA, Ruben SM, Barash SC;
 XX WPI; 2002-665432/71.

XX Novel polypeptide useful for diagnosis, prognosis, prevention, and
 PT treatment of immune, hyperproliferative, renal, respiratory, and
 PT cardiovascular, reproductive, endocrine, gastrointestinal and
 PT neurological disorders.

XX Disclosure; Page 287-288; 335pp; English.

XX The invention relates to an isolated polypeptide comprising a sequence at
 CC least 90% identical to a full length protein sequence selected from 55
 CC sequences given in the specification such as a sequence of 163, 74 or 140
 CC amino acids fully defined in the specification, or the encoding sequence
 CC contained in 49 cDNA clones given in specification e.g. HCFAR05, HWAAB95
 CC or HFNW01. The protein and its encoding nucleic acid are useful for
 CC diagnosing a pathological condition or susceptibility to a pathological
 CC condition in a subject and for preventing, treating or ameliorating a
 CC medical condition. The protein, its encoding nucleic acid and an isolated
 CC antibody that can bind to the protein are useful in treating, preventing,
 CC diagnosing and/or prognosing immunodeficiencies, autoimmune disorders,
 CC allergic reactions and conditions, inflammatory conditions, graft-versus-
 CC host disease, blood-related disorders, hyperproliferative disorders,
 CC renal disorders, cardiovascular disorders, respiratory disorders,
 CC neurological disorders, endocrine disorders, reproductive system
 CC disorders, infectious diseases, and gastrointestinal disorders. The
 CC protein of the invention is useful to stimulate neuronal growth and to
 CC treat, prevent, and/or diagnose neuronal damage which occurs in certain
 CC neuronal disorders or neuro-degenerative conditions, for stimulating
 CC keratinocyte growth, to prevent hair loss, to modulate mammalian
 CC characteristics such as body height, weight, hair color, and to increase
 CC or decrease storage capabilities, fat content, lipid, protein,
 CC carbohydrate, vitamins, minerals, cofactors or other nutritional
 CC components. The nucleic acid of the invention can be used in gene
 CC therapy. This sequence represents a novel human protein of the invention

XX Sequence 238 AA;

Query Match 21.9%; Score 343; DB 5; Length 238;
 Best Local Similarity 33.5%; Pred. No. 3.6e-18;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

QY 34 EHEENPCQLRFLGGTCT---QCTRLCLLFGSGPRCCQCGHGAESDWHLEGGGN 89
 Db 11 ENAHP-CVRAFCAGGSCRKESYDCDCLPLGFLHCXKAITEAI-----56

QY 90 DAP----GOYGFTHDDGELAFPGHVFERSLPVETTELVEVTRSTASGLLWQGVZGE 145
 Db 57 ELPQFTGRSLTYDNPDLKRVSG---SRSS-----NVFMRFKTKADGLLWKG--DSP 105

QY 146 AGQKDFISLGLQDGHVFRYQLGSGEARLNSDEDPINDGWHRVTRALRGRGSIQVDGE 205

Db 106 MRPNSDFISLGLRDSALVFSYNTLGGVASTMWNQSFNDGRWHRVKAIRDGQSGKIIVDDY 165
 QY 206 ELVSGRSPGNVAVNAKSGYVIGGAPDVALITGCRFSSGITGCVKLVLSHARPAPPFQ 265
 Db 166 GARTGKSPGMRLQNLINGALYGVGMKEIALHTRQYMRGLVGCISHFTLST-----DY 218

QY 256 FLDLQHRQAQAGANTRPC 282
 Db 219 HISLVEDAVDGNKINTC 235

RESULT 11
 ABP67037
 ID ABP67037 standard; protein; 238 AA.
 XX ABP67037;
 XX AC ABP67037;
 XX DT 09-DEC-2002 (first entry)
 XX DE Human polypeptide SEQ ID NO 758.

XX Human; noctropic; neuroprotective; cytostatic; dermatological; viroicide;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
 KW antiparkinsonian; anticlacking; antianaemic; antiarthritic; cancer;
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antileucor; anticonvulsant; antifungal;
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.

XX Homo sapiens.
 OS
 XX US2002090672-A1.
 XX 11-JUL-2002.
 XX PF 17-JAN-2001; 2001US-00764853.
 XX PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0190628P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216680P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 14-JUL-2000; 2000US-0217496P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249239P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PA (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX Rosen CA, Ruben SM, Barash SC;
 XX WPI; 2002-681727/73.
 DR N-PSDB; ABV94009.
 XX Novel polypeptide useful for diagnosis, prognosis, prevention, and treatment of immune, hyperproliferative, renal, respiratory, cardiovascular, reproductive, endocrine, gastrointestinal and neurological disorders.
 PT Claim 11; SEQ ID NO 759; 369pp + Sequence Listing; English.
 XX The invention relates to novel genes (ABV83692-ABV84101) and proteins (ABP66710-ABP67129) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

21.9%; Score 343; DB 5; Length 238;
 Best Local Similarity 33.5%; Pred. No. 3.6e-18;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

Qy 34 EHENPQLRPELKHGTC---QVTRCLCLPFGSPRCQGSCHGIAESDWHLEGGSGN 89
 Db 11 ENAAHPC-VRAPCAHGSCKPRKREGYDCDCLPGLFEGELHCKAIIIEAI-----56
 Qy 90 DAP----GQYGAFHDDGFLAFPGHVFSLPEVEFTIELEVRTASGLLLWQGVVEGE 145
 Db 57 EIPQIGRSYLTNDPILKRVSG---SRS-----NVFRPKTKAKDGLLLWRG--DSP 105
 Qy 146 AGQCKDFTSLGLDQGHVPRYQYLGSGEARLVSDPINDGEMHWRTALRGRGSIQVDGE 205
 Db 106 MRFNSDFISLQDRGALVYFSLGSGVSIWVNGSFNDGMRWVRKAVRDQSGKLTVDY 165
 Qy 206 ELVSGRFGFNVAIVAKGYSYIIGAPDVAFLTGGRFSSGITCVKRLVHLSARFGAPPQP 265
 Db 166 GARTKSGFQWMLNIALNGALYVGMKEIALHTNQYMRGLVGCISGHTLST-----DY 218
 Qy 266 PLDLQHRQAQAGANTRPC 282

Db 219 HILSIVEDAVDGNKINTC 235

RESULT 12
 AAU07421
 ID AAU07421 standard; protein; 375 AA.
 XX AAU07421;
 XX 18-DEC-2001 (first entry)
 DE Novel human extracellular matrix (ECM) protein #4.
 KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic; KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological; KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis; KW hyperproliferative disorder; neoplasm; cardiovascular disorder; KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection; KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn; KW wound healing; food additive.
 XX Homo sapiens.
 OS WO200179283-A1.
 XX PN 25-OCT-2001.
 XX PD 11-APR-2001; 2001WO-US011643.
 XX PF 18-APR-2000; 2000US-0198123P.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX Fisciella M, Shi Y, Ebner R, Ruben SM;
 WPI: 2001-611720/70.
 DR N-PSDB; AAS13846.
 XX New nucleic acids encoding extracellular matrix polypeptides, for PT diagnosing, treating, preventing or ameliorating human disorders and PT disease, such as, autoimmune, hyperproliferative or cardiovascular disorders.
 XX Claim 1; Page 297-298; 308pp; English.
 CC The invention relates to novel isolated polynucleotides (I) encoding extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by (I) are used to prevent, treat or ameliorate a medical condition in e.g. CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They CC are also used in diagnosing a pathological condition or susceptibility to CC a pathological condition. The antibodies to the polypeptides can also be CC used in alleviating symptoms associated with the disorders and in CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated CC include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. CC anglogenesis, nervous system disorders e.g. Alzheimer's disease, CC infections caused by bacteria, viruses and fungi and ocular disorders CC e.g. corneal infection. The polypeptides can also be used to aid wound CC healing and epithelial cell proliferation, to prevent skin aging due to CC sunburn, to maintain organs before transplantation, for supporting cell CC culture of primary tissues, to regenerate tissues and in chemotaxis. The CC polypeptides can also be used as a food additive or preservative to CC increase or decrease storage capabilities. The present sequence CC represents the amino acid sequence of novel human extracellular matrix CC (ECM) protein #4
 XX Sequence 375 AA;
 XX Query Match 21.9%; Score 343; DB 4; Length 375;

08-NOV-2000; 2000US-02456610P.
08-NOV-2000; 2000US-02456611P.
08-NOV-2000; 2000US-02456613P.
17-NOV-2000; 2000US-0249207P.
17-NOV-2000; 2000US-0249208P.
17-NOV-2000; 2000US-0249209P.
17-NOV-2000; 2000US-0249210P.
17-NOV-2000; 2000US-0249211P.
17-NOV-2000; 2000US-0249212P.
17-NOV-2000; 2000US-0249213P.
17-NOV-2000; 2000US-0249214P.
17-NOV-2000; 2000US-0249215P.
17-NOV-2000; 2000US-0249216P.
17-NOV-2000; 2000US-0249217P.
17-NOV-2000; 2000US-0249218P.
17-NOV-2000; 2000US-0249244P.
17-NOV-2000; 2000US-0249245P.
17-NOV-2000; 2000US-0249264P.
17-NOV-2000; 2000US-0249265P.
17-NOV-2000; 2000US-0249297P.
17-NOV-2000; 2000US-0249299P.
01-DEC-2000; 2000US-0250160P.
01-DEC-2000; 2000US-0250391P.
05-DEC-2000; 2000US-0251030P.
05-DEC-2000; 2000US-0251988P.
05-DEC-2000; 2000US-0256719P.
06-DEC-2000; 2000US-0251479P.
08-DEC-2000; 2000US-0251856P.
08-DEC-2000; 2000US-0251868P.
08-DEC-2000; 2000US-0251869P.
08-DEC-2000; 2000US-0251989P.
08-DEC-2000; 2000US-0251990P.
11-DEC-2000; 2000US-0254097P.
05-JAN-2001; 2001US-0259678P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX FI
XX PI Rosen CA, Barash SC, Ruben SM;
XX DR
XX DR WPI; 2001-488777/53.
XX DR N-PSDB; AAS28944.
XX PT
XX PT New nucleic acid molecules encoding 49 human secreted proteins for
XX PT diagnosing, preventing, treating or ameliorating medical conditions and
XX PT used as food additives or preservatives.
XX PS
XX PS Claim 11; SEQ ID NO 78; 524pp; English.
XX CC
XX CC The present invention relates to the isolation of novel human uterine
XX CC motility-association polypeptides, and cDNA (AAS28936-AAS28954) and
XX CC genomic sequences encoding for these polypeptides. The sequences of the
XX CC invention are useful in the diagnosis, treatment, prevention and/or
XX CC prognosis of diseases associated with uterine motility such as pregnancy
XX CC and labour, and menstrual disorders. The polynucleotide sequences of the
XX CC invention are also useful in gene therapy. AAU18094-AAU18152 represent
XX CC novel human uterine motility-association polypeptides. Note: The sequence
XX CC data for this patent did not form part of the printed specification, but
XX CC was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX CC
XX SQ Sequence 432 AA;

Query Match 21.9%; Score 343; DB 4; Length 432;
Best Local Similarity 33.5%; Pred. No. 7.1e-18;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
QY 34 EHEENPCQLREPLHGGTC-----CQTRCLCLPFGSPPRCQGSCHGIARSDWHLRSGGN 89
Db 205 ENAAHPC-VRAPCAHGSCRPKREGYDCDFLPGFGLHCKAIIET----- 250
QY 90 DAP-----CQYAYPHDDGFLAPPQHVRSRLPVPVFTELEVRTSTASGLLLMQGVEYCE 145

08-NOV-2000; 2000US-02456610P.
08-NOV-2000; 2000US-02456611P.
08-NOV-2000; 2000US-02456613P.
17-NOV-2000; 2000US-0249207P.
17-NOV-2000; 2000US-0249208P.
17-NOV-2000; 2000US-0249209P.
17-NOV-2000; 2000US-0249210P.
17-NOV-2000; 2000US-0249211P.
17-NOV-2000; 2000US-0249212P.
17-NOV-2000; 2000US-0249213P.
17-NOV-2000; 2000US-0249214P.
17-NOV-2000; 2000US-0249215P.
17-NOV-2000; 2000US-0249216P.
17-NOV-2000; 2000US-0249217P.
17-NOV-2000; 2000US-0249218P.
17-NOV-2000; 2000US-0249244P.
17-NOV-2000; 2000US-0249245P.
17-NOV-2000; 2000US-0249264P.
17-NOV-2000; 2000US-0249265P.
17-NOV-2000; 2000US-0249297P.
17-NOV-2000; 2000US-0249299P.
01-DEC-2000; 2000US-0250160P.
01-DEC-2000; 2000US-0250391P.
05-DEC-2000; 2000US-0251030P.
05-DEC-2000; 2000US-0251988P.
05-DEC-2000; 2000US-0256719P.
06-DEC-2000; 2000US-0251479P.
08-DEC-2000; 2000US-0251856P.
08-DEC-2000; 2000US-0251868P.
08-DEC-2000; 2000US-0251869P.
08-DEC-2000; 2000US-0251989P.
08-DEC-2000; 2000US-0251990P.
11-DEC-2000; 2000US-0254097P.
05-JAN-2001; 2001US-0259678P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX FI
XX PI Rosen CA, Barash SC, Ruben SM;
XX DR
XX DR WPI; 2001-488777/53.
XX DR N-PSDB; AAS28944.
XX PT
XX PT New nucleic acid molecules encoding 49 human secreted proteins for
XX PT diagnosing, preventing, treating or ameliorating medical conditions and
XX PT used as food additives or preservatives.
XX PS
XX PS Claim 11; SEQ ID NO 78; 524pp; English.
XX CC
XX CC The present invention relates to the isolation of novel human uterine
XX CC motility-association polypeptides, and cDNA (AAS28936-AAS28954) and
XX CC genomic sequences encoding for these polypeptides. The sequences of the
XX CC invention are useful in the diagnosis, treatment, prevention and/or
XX CC prognosis of diseases associated with uterine motility such as pregnancy
XX CC and labour, and menstrual disorders. The polynucleotide sequences of the
XX CC invention are also useful in gene therapy. AAU18094-AAU18152 represent
XX CC novel human uterine motility-association polypeptides. Note: The sequence
XX CC data for this patent did not form part of the printed specification, but
XX CC was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX CC
XX SQ Sequence 432 AA;

Query Match 21.9%; Score 343; DB 4; Length 432;
Best Local Similarity 33.5%; Pred. No. 7.1e-18;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
QY 34 EHEENPCQLREPLHGGTC-----CQTRCLCLPFGSPPRCQGSCHGIARSDWHLRSGGN 89
Db 205 ENAAHPC-VRAPCAHGSCRPKREGYDCDFLPGFGLHCKAIIET----- 250
QY 90 DAP-----CQYAYPHDDGFLAPPQHVRSRLPVPVFTELEVRTSTASGLLLMQGVEYCE 145

Db	251	EIQPIGRSYLTNDPDLIKRVSG-----SRS-----NVFMFKTKAKDGLLRG--DSP	299
QY	146	AGGKDFISLQDQCHLVFRYQLSGEARLVSEDPINDGEWHRYVTLAREGRRGSIQVDGE	205
Db	300	MRPNSDFISLGRDGLVFSYMLSGVASINWVGSFNDGRHRVKAVRDQSQSKITVDYD	359
QY	206	ELVSGRSGPENVAIVNAKGSVYIGGAPDVAIITGGRFSSGITGCVKNIILVHSARPPPPQ	265
Db	360	GARTGKSPGMRQLNGALYVGMKEIALHTNRQYMRGLVGCISHFLLST-----DY	412
QY	266	FLDLQHRACAGANTRPC	282
Db	413	HISLVEDAVDGKVINTC	429
RESULT 14			
AAU16938			
ID	AAU16938	standard; protein; 432 AA.	
XX			
AC	AAU16938;		
XX			
DT	07-NOV-2001	(first entry)	
XX			
DE		Human novel secreted protein, SEQ ID 179.	
XX			
KW		Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide; fungicide; opthalmological; vulnesary; secreted protein; rheumatoid arthritis; hyperproliferative disorder; cardiovascular disorder; cardiac arrest; cerebrovascular disorder; cerebral ischaemia; angiogenesis; nervous system disorder; Alzheimer's disease; infection; ocular disorder; corneal infection; wound healing; epithelial cell proliferation; skin ageing; food additive; preservative; antiproliferative.	
OS		Homo sapiens.	
XX			
FN	WC200155441-A2.		
XX			
PD		02-AUG-2001.	
XX			
PF		17-JAN-2001; 2001WO-US001320.	
XX			
PR	31-JAN-2000;	2000US-0179065P.	
PR	04-FEB-2000;	2000US-0180628P.	
PR	24-FEB-2000;	2000US-0184664P.	
PR	02-MAR-2000;	2000US-01866350P.	
PR	16-MAR-2000;	2000US-0189874P.	
PR	17-MAR-2000;	2000US-0190076P.	
PR	18-APR-2000;	2000US-0198123P.	
PR	19-MAY-2000;	2000US-0205515P.	
PR	07-JUN-2000;	2000US-0209467P.	
PR	28-JUN-2000;	2000US-0214866P.	
PR	30-JUN-2000;	2000US-0215135P.	
PR	07-JUL-2000;	2000US-0216647P.	
PR	07-JUL-2000;	2000US-0216880P.	
PR	11-JUL-2000;	2000US-0217487P.	
PR	11-JUL-2000;	2000US-0217496P.	
PR	14-JUL-2000;	2000US-0218290P.	
PR	26-JUL-2000;	2000US-0220963P.	
PR	26-JUL-2000;	2000US-0220964P.	
PR	14-AUG-2000;	2000US-0224518P.	
PR	14-AUG-2000;	2000US-0224519P.	
PR	14-AUG-2000;	2000US-0225213P.	
PR	14-AUG-2000;	2000US-0225214P.	
PR	14-AUG-2000;	2000US-0225266P.	
PR	14-AUG-2000;	2000US-0225267P.	
PR	14-AUG-2000;	2000US-0225268P.	
PR	14-AUG-2000;	2000US-0225270P.	
PR	14-AUG-2000;	2000US-0225447P.	
PR	14-AUG-2000;	2000US-0225477P.	
PR	14-AUG-2000;	2000US-0225758P.	
PR	14-AUG-2000;	2000US-0225759P.	

PR 18-AUG-2000; 2000US-0225279P.
 PR 22-AUG-2000; 2000US-0225681P.
 PR 22-AUG-2000; 2000US-0225686P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 03-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 05-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 25-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235634P.
 PR 27-SEP-2000; 2000US-0235636P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.

PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 06-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251858P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-476222/51.
 N-PSDB; AAS26843.

Novel polypeptides and polynucleotides useful as diagnostic reagents to diagnose diseases or disorders associated with aberrant expression or activity of polypeptides, for treating blood clotting disorder, hemophilia.

Claim 11; SEQ ID NO 179; 601pp; English.

The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present invention represents a novel secreted protein of the invention. Note: The sequence represents a novel secreted protein of the invention. Note: The

Query Match 21.9%; Score 343; DB 4; Length 432;

```

Best Local Similarity 33.5%; Pred. No. 7.1e-18;
Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
Oy 34 EHENPCQLREPLCHGTC---QGRCLCLPQFSPRCCOQSGHGIAESDWHLEGGGN 89
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 205 ENAAHPC-VPAAPCAHGSCRPKEGYDCDPLGFEGLHCOKAIIIAI-----250
Oy 90 DAP-----GQYGAFHDDGFLAFPGHVFPSLSLSEVETFELEVRTASGLLLWQGVVEGE 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 EIPQFIGRSYLTNDPDKLRVSG-----NVFMFKTKAKDGLLWRG--DSP 299
Oy 146 AGOCKDFISLQDGHVFRYQVLSGGEALVSDPINDGEMHRVTLRGRGSIQVNDGE 205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 300 MRPNSDFISLGRDGLVFSYVNLGSGVSIWYNGSFNDGRWRVKAQVDRGSGKITVDDY 359
Oy 206 ELVSGRSPGPNVAVNAGSVYICGAPDVATLIGRFPSSGITCVKQVNLVHSGRPGAPPPQ 265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 GARTGKSPGMRRQLNGALYVGGMKBIALHTNRQYMRGLVGCISHFTLST-----DY 412
Oy 266 FLDLQHRQAQAGANTRPC 282
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 HISLVEDAVDQKNINTC 429

```

RESULT 15

```

ABB10233
ID ABB10233 standard; protein; 432 AA.
XX AC ABB10233;
XX DT 10-JAN-2002 (first entry)
XX XX Human cDNA SEQ ID NO: 541.
DE OS Homo sapiens.
KW Human; Gene therapy; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; inflammation.
XX XX
XX XX
FN WC200154474-A2.
XX XX
PD 02-AUG-2001.
XX XX
PF 17-JAN-2001; 2001WO-US001349.
XX XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205151P.
PR 07-JUN-2000; 2000US-0209467P.
PR 26-JUN-2000; 2000US-0214896P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216680P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0320963P.
PR 26-JUL-2000; 2000US-0320964P.
PR 14-AUG-2000; 2000US-0324518P.
PR 14-AUG-2000; 2000US-0324519P.
PR 14-AUG-2000; 2000US-0325213P.
PR 14-AUG-2000; 2000US-0325214P.
PR 14-AUG-2000; 2000US-0325266P.
PR 14-AUG-2000; 2000US-0325267P.
PR 14-AUG-2000; 2000US-0325268P.
PR 14-AUG-2000; 2000US-0325270P.

```

```

PR 14-AUG-2000; 2000US-0235447P.
PR 14-AUG-2000; 2000US-0235757P.
PR 14-AUG-2000; 2000US-0235758P.
PR 14-AUG-2000; 2000US-0235759P.
PR 18-AUG-2000; 2000US-0236279P.
PR 22-AUG-2000; 2000US-0236681P.
PR 22-AUG-2000; 2000US-0236868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229503P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239933P.
PR 13-OCT-2000; 2000US-0239933P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.

```

Db 300 MRPNSDFISLIGURDGLVFSYNLGSGVASYIMVNGSFNDGRHRVKAVRDGQSGKITVDDY 359
 QY 206 ELVGRSPGPNVAVNAKGSVYIGGAPDVAVITGGRFSSGITGVKNLVLSARPGAPPPQ 265
 Db 360 GARTGKSPGMMROLNINAGLYVGGMKRIALHTNROYMRGLVGCISHFTLST-----DY 412
 QY 266 PLDLQHRQAQAGANTRPC 282
 Db 413 HISLVEDAVDGKNINTC 429

Search completed: March 9, 2004, 17:19:03
 Job time : 35.5822 secs

PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 05-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 05-JAN-2001; 2000US-0254097P.
 PR 05-JAN-2001; 2000US-02549678P.
 XX
 FA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-476161/51.
 DR N-PSDB; ABA06455.
 XX

Isolated nucleic acid molecule encoding an inflammation-associated polypeptide is used in preventing, treating or ameliorating a medical condition.
 Claim 11; SEQ ID NO 541; 859pp + Sequence Listing; English.
 PS The present invention provides human cDNAs, proteins and related genomic XX
 CC DNAs. These can be used in the treatment of neural, immune system, CC
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, CC
 CC renal and proliferative disorders and inflammation. The present sequence CC
 CC is a protein of the invention XX
 SQ Sequence 432 AA;

Query Match 21.9%; Score 343; DB 4; Length 432;
 Best Local Similarity 33.5%; Pred. No. 7.1e-18;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;
 QY 34 EHENPCOLREPCILHGTC-----QETRCILCFGSGPRCQGGHGAESDWHLEGGGN 89
 Db 205 ENAAHFC-VRAPCAHGSRCPRKGYDCDPLGFGHLCKAKATEI----- 250
 QY 90 DAP----GQYGAYFHDGFLAFCYFVRSLEFVETIELEVRSTAGLLWQGEVGE 145
 Db 251 EIPQFTGRSYLTYDNPDIKRVSG-----NVFMRFKTKADGLLLRG--DSP 299
 QY 146 AGGQKDFISLIGLQDGLVFRYQLGSGEARLVSEDPDNGEWHRVTALEGRGSIQVDGE 205

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:11:52 ; Search time 17.0517 Seconds
(without alignments)
2152.832 Million cell updates/sec

Title: US-10-006-011a-3
Perfect score: 3825
Sequence: 1 EIKITRFDPSADGMLYNGQ.....QPLDLQHRAGAGANTREPCPS 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Table with columns: Result No., Score, Match, Length, DB, ID, Description. Contains 45 entries of search results.

Table with columns: ID, PGBM_HUMAN, STANDARD, PRT, 4391_AA. Lists various protein entries and their identifiers.

ALIGNMENTS

RESULT 1
PGBM_HUMAN
ID PGBM_HUMAN STANDARD; PRT; 4391_AA.
AC P98160; Q16287; Q9H3V5;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A. PubMed=1730768;
RX MEDLINE=92112594; PubMed=1730768;
RA Kallunki P., Tryggvason K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor.";
RL J. Cell Biol. 116:559-571(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Colon, and Skin;
RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor.";
RL J. Biol. Chem. 267:8544-8557(1992).
RN [3]
RP SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TYR-1532.
RX MEDLINE=20553141; PubMed=1101850;
RA Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barra D.,
RA Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA Samson D., Urtizbera J.A., Lehmann-Horn F., Weissenbach J.,
RA Hentati F., Fontaine B.;
RT "Perlecan, the major proteoglycan of basement membranes, is altered in
RT patients with Charcot-Marie-Tooth syndrome (Chondrodystrophic myotonia).";
RL Nat. Genet. 26:480-483(2000).
RN [4]
RP SEQUENCE OF 1016-1470 FROM N.A.
RX TISSUE=Colon;
RA Dodge G.R., Kovalszky I., Chu M.L., Haessel J.R., McBride O.W.,
RA Vi H.F., Iozzo R.V.;
RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1.";
RL Genomics 10:673-680(1991).
RN [5]
RP SEQUENCE OF 890-1396 FROM N.A.
RX TISSUE=Fibrosarcoma;

RX MEDLINE=92120660; PubMed=1685141;
 RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
 RA Tytgvaason K.,
 RT "Cloning of human heparan sulfate proteoglycan core protein.
 RT assignment of the gene (HSPG2) to 1p36.1--p25 and identification of
 RT a BamHI restriction fragment length polymorphism.";
 RL Genomics 11:389-396 (1991).
 RN [6]
 RN SEQUENCE OF 1-21 FROM N.A.
 RP MEDLINE=94052171; PubMed=8234307;
 RX Cohen I.R., Graessels S., Murdoch A.D., Iozzo R.V.,
 RA "Structural characterization of the complete human perlecan gene and
 RT its promoter.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408 (1993).
 RN [7]
 RP CARBOHYDRATE-LINKAGE SITE ASN-2121.
 RX MEDLINE=22860472; PubMed=12754519;
 RA Zhang H., Li X.-J., Martin D.B., Abersold R.,
 RT "Identification and quantification of N-linked glycoproteins using
 RT hydrazide chemistry, stable isotope labeling and mass spectrometry.";
 RL Nat. Biotechnol. 21:660-666 (2003).
 CC -!- FUNCTION: This protein is an integral component of basement
 CC membranes. It is responsible for the fixed negative electrostatic
 CC charge and is involved in the charge-selective ultrafiltration
 CC properties. It serves as an attachment substrate for cells.
 CC -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
 CC dimers or stellate structures. It interacts with other basement
 CC membrane components such as laminin, prolargin and collagen type
 CC IV.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: Found in the basement membranes.
 CC -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
 CC AND O-LINKED OLIGOSACCHARIDES.
 CC -!- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
 CC syndrome (SJS1) [MIM:255800]; a rare autosomal recessive disorder
 CC characterized by permanent myotonia (prolonged failure of muscle
 CC relaxation) and skeletal dysplasia, resulting in reduced stature,
 CC kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
 CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
 CC -!- SIMILARITY: Contains 11 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 3 laminin IV domains.
 CC -!- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
 CC -!- SIMILARITY: Contains 3 laminin G-like domains.
 CC -!- SIMILARITY: Contains 4 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 SEA domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X62515; CAA44373.1;
 CC EMBL; M85289; AAA52700.1;
 CC EMBL; AL445795; CAC18534.1;
 CC EMBL; M64283; AAA52699.1;
 CC EMBL; S76436; AAB21121.2;
 CC EMBL; L22078; -; NOT_ANNOTATED_CDS.
 CC EMBL; A38096; A38096.
 CC RSP; P00740; LEDM.
 CC Giens-2DPAGE; P98160; -;
 CC Genew; HGNC:5273; HSPG2.
 CC MIM; 142461; -;
 CC MIM; 255800; -;
 CC InterPro; IPR008985; ConA_like_lec_gl.
 CC InterPro; IPR00742; EGF_2.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR006210; IEGF.
 CC InterPro; IPR007110; IG-like.
 CC InterPro; IPR003599; IG.
 CC InterPro; IPR003598; Ig_c2.

DR	InterPro; IPR003596; Ig_v.	22	4391
DR	InterPro; IPR000034; Laminin_B.	80	194
DR	InterPro; IPR002049; Laminin_EGF.	198	235
DR	InterPro; IPR001791; Laminin_G.	284	320
DR	InterPro; IPR002172; LDL_receptor_A.	324	360
DR	InterPro; IPR000082; SEA_domain.	367	404
DR	Pfam; PF00008; EGF; 4.	405	504
DR	Pfam; PF00047; IG; 22.	521	530
DR	Pfam; PF00052; laminin_B; 3.	531	730
DR	Pfam; PF00053; laminin_EGF; 7.	731	763
DR	Pfam; PF00054; laminin_G; 3.	764	813
DR	Pfam; PF00057; ldl_recept_a; 4.	814	871
DR	Pfam; PF01390; SEA; 1.	879	933
DR	PRINTS; PR00261; LDLRECEPTOR.	924	933
DR	ProDom; PD003031; Laminin_B; 3.	934	933
DR	SMART; SM00181; EGF; 15.	924	934
DR	SMART; SM00180; EGF_Lam; 12.	1126	1158
DR	SMART; SM00409; IG; 22.	1159	1208
DR	SMART; SM00408; IGC2; 21.	1209	1265
DR	SMART; SM00406; IGV; 7.	1275	1324
DR	SMART; SM00281; Lamb; 3.	1325	1334
DR	SMART; SM00282; Lamg; 3.	1335	1529
DR	SMART; SM00192; LDLA; 4.	1530	1562
DR	SMART; SM00200; SEA; 1.	1563	1612
DR	PROSITE; PS00022; EGF_1; 9.	1613	1670
DR	PROSITE; PS01186; EGF_2; 6.	1672	1771
DR	PROSITE; PS50026; EGF_3; 4.	1772	1865
DR	PROSITE; PS50835; IG_LIKE; 22.	1866	1955
DR	PROSITE; PS50025; LAM_G_DOMAIN; 3.	1956	2051
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 11.	2052	2151
DR	PROSITE; PS01209; LDLRA_1; 4.	2152	2244
DR	PROSITE; PS50068; LDLRA_2; 4.	2245	2340
DR	PROSITE; PS50024; SEA; 1.	2341	2436
KW	Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;	2437	2533
KW	Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;	2534	2629
KW	Extracellular matrix; EGF-like domain; Disease mutation.	2630	2726
FT	POTENTIAL.	1	21
FT	SIGNAL	22	4391
FT	CHAIN	80	194
FT	DOMAIN	198	235
FT	LDL-RECEPTOR CLASS A 1.	284	320
FT	LDL-RECEPTOR CLASS A 2.	324	360
FT	LDL-RECEPTOR CLASS A 3.	367	404
FT	LDL-RECEPTOR CLASS A 4.	405	504
FT	IG-LIKE C2-TYPE 1.	521	530
FT	LAMININ EGF-LIKE 1 (N-TERMINAL).	531	730
FT	LAMININ DOMAIN IV 1 (DOMAIN III A).	731	763
FT	LAMININ EGF-LIKE 1 (C-TERMINAL).	764	813
FT	LAMININ EGF-LIKE 2.	814	871
FT	LAMININ EGF-LIKE 3.	879	933
FT	LAMININ EGF-LIKE 4 (INCOMPLETE).	924	933
FT	LAMININ EGF-LIKE 5 (N-TERMINAL).	934	933
FT	LAMININ DOMAIN IV 2 (DOMAIN III B).	1126	1158
FT	LAMININ EGF-LIKE 5 (C-TERMINAL).	1159	1208
FT	LAMININ EGF-LIKE 6.	1209	1265
FT	LAMININ EGF-LIKE 7.	1275	1324
FT	LAMININ EGF-LIKE 8.	1325	1334
FT	LAMININ EGF-LIKE 9 (N-TERMINAL).	1335	1529
FT	LAMININ DOMAIN IV 3 (DOMAIN III C).	1530	1562
FT	LAMININ EGF-LIKE 9 (C-TERMINAL).	1563	1612
FT	LAMININ EGF-LIKE 10.	1613	1670
FT	LAMININ EGF-LIKE 11.	1672	1771
FT	IG-LIKE C2-TYPE 2.	1772	1865
FT	IG-LIKE C2-TYPE 3.	1866	1955
FT	IG-LIKE C2-TYPE 4.	1956	2051
FT	IG-LIKE C2-TYPE 5.	2052	2151
FT	IG-LIKE C2-TYPE 6.	2152	2244
FT	IG-LIKE C2-TYPE 7.	2245	2340
FT	IG-LIKE C2-TYPE 8.	2341	2436
FT	IG-LIKE C2-TYPE 9.	2437	2533
FT	IG-LIKE C2-TYPE 10.	2534	2629
FT	IG-LIKE C2-TYPE 11.	2630	2726
FT	IG-LIKE C2-TYPE 12.	2630	2726

Qy	Db	Query Match	100.0%	Score	3825	DB 1	Length	4391	Matches	705	Conservative	0	Mismatches	0	Indels	0	Gaps	0
1	EIKITRPSADGMLYNGQKVPSPPTLNANRQPDFISFLVGGRRPEFRFDAGSGWATI	60																
3687	EIKITRPSADGMLYNGQKVPSPPTLNANRQPDFISFLVGGRRPEFRFDAGSGWATI	3746																
61	RHPTPLALGHFHTVLLRSLTQSSIVGLDLPVNTSQKFGDLNBEELYGYPPYGA	120																
3747	RHPTPLALGHFHTVLLRSLTQSSIVGLDLPVNTSQKFGDLNBEELYGYPPYGA	3806																
121	IPKAGLSGPGVRELRIOGSEIVPHDLNLAHGHSHCPTCRDRPCONGQCHDSRSSS	180																
3807	IPKAGLSGPGVRELRIOGSEIVPHDLNLAHGHSHCPTCRDRPCONGQCHDSRSSS	3866																
181	YVCVPAAGTGRSCEHSQALHCHPERACGPDATCVNRPDGRGVTCKHLGRSGLRCESGVT	240																
3867	YVCVPAAGTGRSCEHSQALHCHPERACGPDATCVNRPDGRGVTCKHLGRSGLRCESGVT	3926																
241	VTPPSLGSAGSYLALPALTWHHELRLDVEFKPLADPGVLLFSGGKSPVEDFVSLAMVG	300																
3927	VTPPSLGSAGSYLALPALTWHHELRLDVEFKPLADPGVLLFSGGKSPVEDFVSLAMVG	3986																
301	GHLEFYEYELGSLAVLRSAPALGRMHRVSAERLNKDGSLRVNGRVPVLRSSPKGSQGL	360																
3987	GHLEFYEYELGSLAVLRSAPALGRMHRVSAERLNKDGSLRVNGRVPVLRSSPKGSQGL	4046																
361	NLHTLLYLVGVEVPSPLSPATNWSAHFRFCVGEVSVNGKRLDLYSFLGSGIGQCYDSS	420																
4047	NLHTLLYLVGVEVPSPLSPATNWSAHFRFCVGEVSVNGKRLDLYSFLGSGIGQCYDSS	4106																
421	PCERQPCQAGTAMPAGEYFQCLRDGPKGDLCSHEENPCOLRPPCLHGTCQGRCLC	480																
4107	PCERQPCQAGTAMPAGEYFQCLRDGPKGDLCSHEENPCOLRPPCLHGTCQGRCLC	4166																
481	LPFGSPRCQSGHGIASDWHLEGGNDAPGQVGFYFDDGFLAPFGHVFVRSIPEV	540																
4167	LPFGSPRCQSGHGIASDWHLEGGNDAPGQVGFYFDDGFLAPFGHVFVRSIPEV	4226																
541	PEFIEVTRTASGLLWQVEVAGQKDFISLGLDGHVFRYQLSGEARLVSED	600																
4227	PEFIEVTRTASGLLWQVEVAGQKDFISLGLDGHVFRYQLSGEARLVSED	4286																
601	PINDGEWHRVTALREGRRSIOVDGSELYSGRSPGNVAVNAKGSVYIGGAPDVATLTGG	660																
4287	PINDGEWHRVTALREGRRSIOVDGSELYSGRSPGNVAVNAKGSVYIGGAPDVATLTGG	4346																
661	RFSSGITGVKNIYLSRSPGPPPPQPLDQHRAGANTRPCPS	705																
4347	RFSSGITGVKNIYLSRSPGPPPPQPLDQHRAGANTRPCPS	4391																

RESULT 2
 ID_PGBM_MOUSE STANDARD; PRT: 3707 AA.
 AC Q05793;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Basement membrane-specific heparan sulfate proteoglycan core
 DE protein precursor (HSPG) (Perlecan) (PLC).
 GN Hspg2
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Melanoma;
 RX MEDLINE=92078153; PubMed=1744087;
 RA Noonan D.M., Fuller A., Valente P., Cai S., Horigan E., Sasaki M.,

Yamada Y., Hassell J.R.;
 "The complete sequence of perlecan, a basement membrane heparan sulfate proteoglycan, reveals extensive similarity with laminin A chain, low density lipoprotein-receptor, and the neural cell adhesion molecule";
 J. Biol. Chem. 266:22939-22947(1991).
 [2]
 SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
 MEDLINE=89034110; PubMed=2972708;
 RX Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M., Yamada Y., Hassell J.R.;
 "Identification of cDNA clones encoding different domains of the basement membrane heparan sulfate proteoglycan";
 J. Biol. Chem. 263:16379-16387(1988).
 CC -!- FUNCTION: This protein is an integral component of basement membranes. It is responsible for the fixed negative electrostatic charge and is involved in the charge-selective ultrafiltration properties. It serves as an attachment substrate for cells.
 CC -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in dimers or stellate structures. It interacts with other basement membrane components such as laminin, prolargin and collagen type IV.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: Found in the basement membranes.
 CC -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
 CC -!- SIMILARITY: Contains 11 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 3 laminin IV domains.
 CC -!- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
 CC -!- SIMILARITY: Contains 3 laminin G-like domains.
 CC -!- SIMILARITY: Contains 1 EGF-like domain.
 CC -!- SIMILARITY: Contains 1 SEA domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

 EMBL: W77174; AAA39911.1; -
 EMBL: J04054; AAA39899.1; -
 EMBL: J04055; AAA39912.1; -
 DR PIR; S18252; S18252.
 DR PDB; 1GL4; 28-NOV-01.
 DR MGD; MGI:96257; Hspg2.
 DR GO; GO:0005604; C:basement membrane; IDA.
 DR InterPro; IPR008985; Cona_like lec_gl.
 DR InterPro; IPR00742; EGF 2.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR007110; I9-like.
 DR InterPro; IPR003598; I9 C2.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR002172; Ldl_receptor_A.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF00007; I9; 4.
 DR Pfam; PF00052; laminin_B; 3.
 DR Pfam; PF00053; laminin_EGF; 7.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF00057; ldl_recept_a; 4.
 DR Pfam; PF01390; SEA; 1.
 DR PRINTS; PRO0261; LDLRECEPTOR.
 DR ProDom; PD003031; Laminin_B; 3.
 DR SMART; SM00180; EGF_Lam; 7.
 DR SMART; SM00408; I9C2; 14.
 DR SMART; SM00281; LamB; 3.
 DR SMART; SM00282; LamG; 3.

DR SMART; SM00192; LDLA; 4.
 DR SMART; SM00200; SEA; 1.
 DR PROSITE; PS00022; EGF_1; 8.
 DR PROSITE; PS01186; EGF_2; 5.
 DR PROSITE; PS00266; EGF_3; 4.
 DR PROSITE; PS50835; IG_LIKE; 15.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 DR PROSITE; PS01209; LDLRA_1; 4.
 DR PROSITE; PS50068; LDLRA_2; 4.
 DR PROSITE; PS50024; SEA; 1.
 KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
 KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
 KW Extracellular matrix; EGF-like domain; 3D-structure.
 FT SIGNAL 1 21
 FT CHAIN 22 3707
 FT DOMAIN 80 194
 FT DOMAIN 195 234
 FT DOMAIN 281 319
 FT DOMAIN 320 359
 FT DOMAIN 360 403
 FT DOMAIN 404 504
 FT DOMAIN 521 530
 FT DOMAIN 531 730
 FT DOMAIN 731 763
 FT DOMAIN 764 813
 FT DOMAIN 814 871
 FT DOMAIN 879 923
 FT DOMAIN 924 933
 FT DOMAIN 1126 1158
 FT DOMAIN 1159 1208
 FT DOMAIN 1209 1265
 FT DOMAIN 1275 1324
 FT DOMAIN 1325 1334
 FT DOMAIN 1335 1529
 FT DOMAIN 1530 1562
 FT DOMAIN 1563 1612
 FT DOMAIN 1613 1670
 FT DOMAIN 1671 1771
 FT DOMAIN 1772 1865
 FT DOMAIN 1866 1954
 FT DOMAIN 1955 2049
 FT DOMAIN 2050 2148
 FT DOMAIN 2149 2244
 FT DOMAIN 2245 2343
 FT DOMAIN 2344 2436
 FT DOMAIN 2437 2532
 FT DOMAIN 2533 2619
 FT DOMAIN 2620 2720
 FT DOMAIN 2721 2809
 FT DOMAIN 2810 2895
 FT DOMAIN 2896 2980
 FT DOMAIN 2984 3162
 FT DOMAIN 3163 3241
 FT DOMAIN 3245 3425
 FT DOMAIN 3518 3705
 FT SITE 65 67
 FT SITE 71 73
 FT SITE 76 78
 FT SITE 3617 3617
 FT DISULFID 199 212
 FT DISULFID 206 225
 FT DISULFID 219 234
 FT DISULFID 285 297
 FT DISULFID 292 310
 FT DISULFID 304 319
 FT DISULFID 325 337
 FT DISULFID 332 350
 FT DISULFID 344 359
 FT DISULFID 368 381

FT DISULFID 375 394
 FT DISULFID 388 403
 FT DISULFID 428 479
 FT DISULFID 764 773
 FT DISULFID 766 780
 FT DISULFID 783 792
 FT DISULFID 795 811
 FT DISULFID 814 829
 FT DISULFID 816 839
 FT DISULFID 842 851
 FT DISULFID 854 869
 FT DISULFID 1159 1168
 FT DISULFID 1161 1175
 FT DISULFID 1178 1187
 FT DISULFID 1190 1206
 FT DISULFID 1209 1224
 FT DISULFID 1211 1234
 FT DISULFID 1237 1246
 FT DISULFID 1249 1263
 FT DISULFID 1275 1287
 FT DISULFID 1277 1293
 FT DISULFID 1295 1304
 FT DISULFID 1307 1322
 FT DISULFID 1563 1572
 FT DISULFID 1565 1579
 FT DISULFID 1582 1591
 FT DISULFID 1594 1610
 FT DISULFID 1613 1628
 FT DISULFID 1641 1650
 FT DISULFID 1653 1666
 FT DISULFID 1792 1839
 FT DISULFID 1866 1932
 FT DISULFID 1976 2021
 FT DISULFID 2073 2118
 FT DISULFID 2170 2215
 FT DISULFID 2268 2313

Query Match 88.6%; Score 3389.5; DB 1; Length 3707;
 Best Local Similarity 88.9%; Pred. No. 3.6e-206;
 Matches 627; Conservative 33; Mismatches 40; Indels 5; Gaps 3;

QY 1 EKIIFRSDSADGMLLYNQKRVFSGPTNLANRQPDFISFLVGRPFERFDAGSGMATI 60
 3008 EKIIFRSDSADGMLLYNQKRVFSGPTNLANRQPDFISFLVGRPFERFDAGSGMATI 3064
 QY 61 REPTPLALCHFHVTLLRSITQGLIYVGDLPVNGTSGKFOGLDNEELVLYGYPDYGA 120
 3065 REPTPLALCHFHVTLLRSITQGLIYVGDLPVNGTSGKFOGLDNEELVLYGYPDYGA 3124
 QY 121 IPKAGLSSGFTGCVRELRIOGSEIVFHDNLNTHANGISHCFTCRDRPCQMGCCQCHDSSESS 180
 3125 IPKAGLSSGFTGCVRELRIOGSEIVFHDNLNTHANGISHCFTCRDRPCQMGCCQCHDSSESS 3184
 QY 181 YVVCVCPAGFTGSRCEHSQALHCHPACGPPATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
 3185 YVVCVCPAGFTGSRCEHSQALHCHPACGPPATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 3243
 QY 241 VYTPSLSGANGYLALPALTHTHRLDVEFKPLADPCVLLFSGKSGPVYEDVFLAMVG 300
 3244 VYTPSLSGANGYLALPALTHTHRLDVEFKPLADPCVLLFSGKSGPVYEDVFLAMVG 3303
 QY 301 GHLEFRYELGSLAVLRSAAEPLALGRWHVSAERLNKDGSLRVNGRSPVLRSSPFGSKS 360
 3304 GHLEFRYELGSLAVLRSAAEPLALGRWHVSAERLNKDGSLRVNGRSPVLRSSPFGSKS 3363
 QY 361 NLHTLLYLGVEPSPVPLSPATMSAHFGCVGVSVNGKRLDLYTSLPQSGQYGCQYDSS 420
 3364 NLHTLLYLGVEPSPVPLSPATMSAHFGCVGVSVNGKRLDLYTSLPQSGQYGCQYDSS 3423
 QY 421 PCERQPCGHGATCMPAGEYEFQCLCRDGFKGDLCHEHENPCQLRHGTCGCTRC 480
 3424 PCERQPCGHGATCMPAGEYEFQCLCRDGFKGDLCHEHENPCQLRHGTCGCTRC 3483

QY 481 LPFSGRCQGGHGIASDWHLEGGNDAPGQYGFYFDHDDGFLAPFFGVFRSLPEV 540
 DB 3484 LPFSGRCQGGAGYGVESDWHPEGGNDAPGQYGFYFDHDDGFLAPFFGVFRSLPEV 3543
 QY 541 PETIELEVRSTASGLLQGVVEGAGOGKDFISLGLQDHLVFRYQLGSGEARLYSED 600
 DB 3544 PETIELEVRSTADGLLQGVV-VRASRSKDFISLGLQDHLVFRYQLGSGEARLYSGD 3602
 QY 601 PINDGEHVRVTLREGRRSIQVDGEEELVSGRSPGNVAVNAKGSVYIGGAPDVATLTTGG 660
 DB 3603 PINDGEHVRVTLREGRRSIQVDGEEELVSGRSPGNVAVNAKGSVYIGGAPDVATLTTGG 3662
 QY 661 RFSGGITGCVKLVLSRAGPAPPPOPLDQHRAGAGANTRECPSS 705
 DB 3663 KFSSGITGCVKLVLSRAGPAPPPOPLDQHRAGAGANTRECPSS 3707

RESULT 3
 AGRI RAT STANDARD; PRT; 1959 AA.
 AC P25304; Q63034;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Agrin precursor.
 GN AGRN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=101116;
 RN [1]
 RP SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
 RC TISSUE=Embryonic spinal cord;
 RX MEDLINE=91222570; PubMed=181019;
 RA Rupp F., Payan D.G., Magill-Soic C., Cowan D.M., Scheller R.H.;
 RT "Structure and expression of a rat agrin.";
 RL Neuron 6:811-823(1991).
 RN [2]
 RP SEQUENCE OF 1777-1801 FROM N.A.
 RX MEDLINE=92407628; PubMed=1326608;
 RA Rupp F., Czeceklik T., Linnal M., Peterson K., Francke U., Scheller R.;
 RT "Structure and chromosomal localization of the mammalian agrin gene.";
 RL J. Neurosci. 12:3535-3544(1992).
 CC -!- FUNCTION: Component of the basal lamina that causes the
 aggregation of acetylcholine receptors and acetylcholine-esterase
 on the surface of muscle fibers of the neuromuscular junction.
 CC -!- SUBUNIT: Binds to laminin.
 CC -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
 junction.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=5;
 CC Comment=Additional isoforms seem to exist. Isoforms differ in
 their acetylcholine receptor clustering activity;
 CC Name=1;
 CC IsoId=P25304-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P25304-2; Sequence=VSP_001365;
 CC Name=3;
 CC IsoId=P25304-3; Sequence=VSP_001366;
 CC Name=4;
 CC IsoId=P25304-4; Sequence=VSP_001367;
 CC Name=5;
 CC IsoId=P25304-5; Sequence=VSP_001368;
 CC -!- TISSUE SPECIFICITY: Embryonic nervous system and muscle.
 CC -!- DEVELOPMENTAL STAGE: More abundant early in development.
 CC -!- PTM: Contains heparan sulfate chains as well as N-linked and O-
 linked oligosaccharides (By similarity).
 CC -!- SIMILARITY: Contains 9 Kazal-like domains.
 CC -!- SIMILARITY: Contains 2 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 4 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 SEA domain.
 CC -!- SIMILARITY: Contains 3 laminin G-like domains.

CC -!- CAUTION: It is uncertain whether Met-1, Met-18 or Met-24 is the
 initiator.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M64780; AAA40702.1; ALT_INIT.
 CC EMBL; S44194; AAB23326.1; .
 CC EIR; JH0389; AGRT
 CC HSP; P00740; IEDM.
 CC InterPro; IPR008985; ConA_like_1ec_gl.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR003645; FcIN.
 CC InterPro; IPR002350; Kazal.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR001791; Laminin_G.
 CC Pfam; PF00008; EGF; 4.
 CC Pfam; PF00050; Kazal; 9.
 CC Pfam; PF00053; Laminin_EGF; 2.
 CC Pfam; PF00054; Laminin_G; 3.
 CC Pfam; PF01390; SEA; 1.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC SMART; SM00180; EGF_Lam; 2.
 CC SMART; SM00274; FcIN; 8.
 CC SMART; SM00280; KAZAL; 9.
 CC SMART; SM00282; LamG; 3.
 CC SMART; SM00200; SEA; 1.
 CC PROSITE; PS00022; EGF_1; 6.
 CC PROSITE; PS01186; EGF_2; 1.
 CC PROSITE; PS50026; EGF_3; 4.
 CC PROSITE; PS50025; LAM_G_DOMAIN; 3.
 CC PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 CC GlycoProtId; EGF-like domain; Repeat; Alternative splicing; Signal;
 CC Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
 CC SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 1959 AGRIN.
 FT DOMAIN 65 137 KAZAL-LIKE 1.
 FT DOMAIN 141 212 KAZAL-LIKE 2.
 FT DOMAIN 213 284 KAZAL-LIKE 3.
 FT DOMAIN 287 356 KAZAL-LIKE 4.
 FT DOMAIN 361 429 KAZAL-LIKE 5.
 FT DOMAIN 430 494 KAZAL-LIKE 6.
 FT DOMAIN 495 559 KAZAL-LIKE 7.
 FT DOMAIN 563 645 KAZAL-LIKE 8.
 FT DOMAIN 688 741 LAMININ EGF-LIKE 1.
 FT DOMAIN 742 788 LAMININ EGF-LIKE 2.
 FT DOMAIN 794 864 KAZAL-LIKE 9.
 FT DOMAIN 1023 1145 SEA.
 FT DOMAIN 1220 1258 EGF-LIKE 1.
 FT DOMAIN 1263 1439 LAMININ G-LIKE 1.
 FT DOMAIN 1440 1477 EGF-LIKE 2.
 FT DOMAIN 1479 1516 EGF-LIKE 3.
 FT DOMAIN 1526 1708 LAMININ G-LIKE 2.
 FT DOMAIN 1709 1748 EGF-LIKE 4.
 FT DOMAIN 1784 1956 LAMININ G-LIKE 3.
 FT DOMAIN 869 992 SER/THR-RICH.
 FT DOMAIN 1147 1215 SER/THR-RICH.
 FT DISULFID 97 116 POTENTIAL.
 FT DISULFID 105 137 POTENTIAL.
 FT DISULFID 171 191 POTENTIAL.
 FT DISULFID 180 212 POTENTIAL.
 FT DISULFID 244 263 POTENTIAL.
 FT DISULFID 252 284 POTENTIAL.
 FT DISULFID 316 335 POTENTIAL.
 FT DISULFID 324 356 POTENTIAL.

FT	DISULFID	389	408	POTENTIAL.	339	GSRLVNGRPLVRSRSPGKSOG	---	LNHLTLLYLGGVPSVPLSPATNMSAHRGCVGE	393
FT	DISULFID	397	429	POTENTIAL.	1627	GALQVGDGPRVLSGESP	---	KSRVPHMLNKEFLYIGGAPDFSKLARGAVSSGFGVIQL	1685
FT	DISULFID	454	473	POTENTIAL.					
FT	DISULFID	462	494	POTENTIAL.					
FT	DISULFID	518	538	POTENTIAL.	394	VSVNGKRLDLYTFLGSGOIGCQDSSPCERQPCOHGATCMPAGEYEFQCLCRDGFKGD	---	L	453
FT	DISULFID	527	559	POTENTIAL.	1686	VSLRGHQL	---	LTQEHVLRVAVDVSFADHPCTQA	1716
FT	DISULFID	604	624	POTENTIAL.					
FT	DISULFID	613	645	POTENTIAL.	454	CEHERNPPCLREFCLHGQTC	---	QOT-RCLCLPFGSPRCQOQSGHGHTARSDBWHLESGG	509
FT	DISULFID	688	700	BY SIMILARITY.					
FT	DISULFID	690	707	BY SIMILARITY.					
FT	DISULFID	709	718	BY SIMILARITY.					
FT	DISULFID	721	739	BY SIMILARITY.					
FT	DISULFID	742	754	BY SIMILARITY.					
FT	DISULFID	744	761	BY SIMILARITY.					
FT	DISULFID	763	772	BY SIMILARITY.					
FT	DISULFID	775	786	BY SIMILARITY.					
FT	DISULFID	823	843	POTENTIAL.					
FT	DISULFID	832	864	POTENTIAL.					
FT	DISULFID	1224	1235	BY SIMILARITY.					
FT	DISULFID	1229	1246	BY SIMILARITY.					
FT	DISULFID	1248	1257	BY SIMILARITY.					
FT	DISULFID	1444	1455	POTENTIAL.					
FT	DISULFID	1449	1465	POTENTIAL.					
FT	DISULFID	1467	1476	POTENTIAL.					
FT	DISULFID	1483	1494	BY SIMILARITY.					
FT	DISULFID	1488	1504	BY SIMILARITY.					
FT	DISULFID	1506	1515	BY SIMILARITY.					
FT	DISULFID	1713	1727	BY SIMILARITY.					
FT	DISULFID	1721	1736	BY SIMILARITY.					
FT	DISULFID	1738	1747	BY SIMILARITY.					
FT	CARBOHYD	145	145	N-LINKED (GLCNAC. . .)					
FT	CARBOHYD	572	672	N-LINKED (GLCNAC. . .)					
FT	CARBOHYD	827	827	N-LINKED (GLCNAC. . .)					
FT	CARBOHYD	957	957	N-LINKED (GLCNAC. . .)					
FT	VARSPLIC	1144	1152	Missing (in isoform 2).					
FT	VARSPLIC	1780	1798	Missing (in isoform 3).					
FT	VARSPLIC	1788	1798	Missing (in isoform 4).					
FT	VARSPLIC	1780	1787	/FtId=VSP_001367.					
FT	VARSPLIC	314	314	Missing (in isoform 5).					
FT	VARIANT	1959	AA; 208645	W; 7FEEDFAFF89CC31					
SQ	SEQUENCE								
	Query Match		22.5%	Score 859; DB 1; Length 1959;					
	Best Local Similarity		31.5%	prod No 1.4e-46;					
	Matches	241;	Conservative	92; Mismatches 278; Indels 154; Gaps 23;					
Qy	2	IKITFRPDSADGMLLYNQKRVPGSPPTNLANRQPDFISFLVGGPRPFRPADGSGMATIR	61						
Db	1287	LALFRALETEGLLYNGNA	-----	RGKDFLALALLDGRVQFRFDTGGSPAVLT	1335				
Qy	62	HPTPLALGHFTVTLRLSQTGSLIYDGLAPVNGTSQKFGQLDLNLSLYLGGVDPYG	- 119						
Db	1336	SLVPEGRHRLSLSRHMQTSLVDGETPVGESPSCTDGLNLDNLNLYVGGIPBEQVA	1395						
Qy	120	-ALPKAGLSGFGTGVRELRIOEEIVFDLNLTA	---	HGISHC--PTCRDRPCQNGQC	173				
Db	1396	MVLDRTSVGVGLKGCINMLDINNQQLELSDWRAAVQSSGVGECGDHPCLNPHCGGALC	1455						
Qy	174	HDSRSSYVVCVAGFTGSRCEHSQALHCHPCACGPDATCVNRPDGRVTCRCHLGRSL	233						
Db	1456	QALEAGMFLCQPPRGFPFCADKES	- PCQPNPCHGAAPCVLSSG-GAKCECPGRSGT	1513					
Qy	234	RCEGVVTTTPTSLSGAG	-----	SYLALPALTNTHHEL	---	RLDVFEPKLPDGP	278		
Db	1514	FCQ	---	LETAGSRPFLADFNFGYSLELKLHFTFERDLGKMALEVMFLARGPSSG	1566				
Qy	279	VLFPSGKGSFVEDFVSLAVMGHLEFRYELGSLVLRSAEPLALGRWHRVSAERLNKD	338						
Db	1567	LLLYNGKTKDGKDFVSLALHNLHLEFFCYDLGKGAIVRSKEPIALGTWVRVFLERNGRK	1626						

RESULT 4

AGRI_CHICK	STANDARD;	PRT; 1955	AA.
ID	AGRI_CHICK		
AC	P31696;		
DT	01-JUL-1993	(Rel. 26, Created)	
DT	01-JUL-1993	(Rel. 26, Last sequence update)	
DT	28-FEB-2003	(Rel. 41, Last annotation update)	
DE	Agrin precursor.		
GN	AGRN.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
CC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RY	MEDLINE=92232297; PubMed=1314620;		
RA	Ts'ao K.W.K., Ruegg M.A., Escher G., Kroeger S., McMahan U.J.;		
RT	"cDNA that encodes active agrin.";		
RL	Neuron 8:677-689(1992).		
RN	[2]		
RP	ALTERNATIVE SPLICING.		
RX	MEDLINE=92232298; PubMed=1314621;		
RA	Ruegg M.A., Ts'ao K.W.K., Horton S.E., Kroeger S., Escher G.,		
RA	Gensch E.M., McMahan U.J.;		
RT	"The agrin gene codes for a family of basal lamina proteins that		
RT	differ in function and distribution.";		
RL	Neuron 8:691-699(1992).		
CC	-1- FUNCTION: Component of the basal lamina that causes the		
CC	aggregation of acetylcholine receptors and acetylcholine-esterase		
CC	on the surface of muscle fibers of the neuromuscular junction.		
CC	-1- SUBUNIT: Binds to laminin.		
CC	-1- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular		
CC	junction.		
CC	-1- ALTERNATIVE PRODUCTS:		
CC	Event=Alternative splicing; Named isoforms=3;		
CC	Comment=Additional isoforms seem to exist. Isoforms differ in		
CC	their acetylcholine receptor clustering activity;		
CC	Name=1;		
CC	ISOID=P31696-1; Sequence=Displayed;		
CC	Name=2; Synonyms=Agrin-related protein 1;		
CC	ISOID=P31696-2; Sequence=VSP_001370;		
CC	Name=3; Synonyms=Agrin-related protein 2;		

CC IsoId=P31696-3; Sequence=VSP_001369, VSP_001370;
 CC -!- Contains heparan sulfate chains as well as N-linked and O-
 CC linked oligosaccharides (By similarity).
 CC -!- SIMILARITY: Contains 9 Kazal-like domains.
 CC -!- SIMILARITY: Contains 2 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 4 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 SEA domain.
 CC -!- SIMILARITY: Contains 3 laminin G-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----

DR EMBL; M94271; AAA48585.1; -;
 DR EMBL; M97371; AAA48586.1; -;
 DR PIR; JH0591; AGCH.
 DR HSSP; P00740; LEDM.
 DR InterPro; IPR004850; Agrin Nta.
 DR InterPro; IPR000152; Asx_hydroxyl_s.
 DR InterPro; IPR008985; Cona1_like_lec_gl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR003645; FOLN.
 DR InterPro; IPR002350; kazal.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR000082; SEA_domain.
 DR InterPro; IPR008993; TIME_like.
 DR Pfam; PF00008; EGF_4.
 DR Pfam; PF00050; kazal; 9
 DR Pfam; PF00053; laminin_EGF; 2.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF0146; NTA; 1.
 DR Pfam; PF01390; SEA; 1.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00180; EGF_Lam; 2.
 DR SMART; SM00274; FOLN; 8.
 DR SMART; SM00280; KAZAL; 9.
 DR SMART; SM00282; LamG; 3.
 DR SMART; SM00200; SEA; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; 6.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50026; EGF_3; 4.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 DR PROSITE; PS50024; SEA; 1.
 KW Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
 KW Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
 FT SIGNAL 1 38 POTENTIAL.
 FT CHAIN 39 1955 AGRIN.
 FT DOMAIN 54 126 KAZAL-LIKE 1.
 FT DOMAIN 130 201 KAZAL-LIKE 2.
 FT DOMAIN 202 273 KAZAL-LIKE 3.
 FT DOMAIN 276 344 KAZAL-LIKE 4.
 FT DOMAIN 350 418 KAZAL-LIKE 5.
 FT DOMAIN 419 483 KAZAL-LIKE 6.
 FT DOMAIN 484 548 KAZAL-LIKE 7.
 FT DOMAIN 551 633 KAZAL-LIKE 8.
 FT DOMAIN 675 728 LAMININ_EGF-LIKE 1.
 FT DOMAIN 729 775 LAMININ_EGF-LIKE 2.
 FT DOMAIN 801 851 KAZAL-LIKE 9.
 FT DOMAIN 1025 1148 SEA.
 FT DOMAIN 1229 1265 EGF-LIKE 1.
 FT DOMAIN 1270 1445 LAMININ_G-LIKE 1.
 FT DOMAIN 1446 1483 EGF-LIKE 2.
 FT DOMAIN 1485 1522 EGF-LIKE 3.
 FT DOMAIN 1532 1713 LAMININ_G-LIKE 2.

DOMAIN	1714	1752	EGF-LIKE 4.
FT DOMAIN	1776	1952	LAMININ G-LIKE 3.
FT DOMAIN	856	995	SER/THR-RICH.
FT DOMAIN	1150	1219	SER/THR-RICH.
FT DISULFID	86	105	POTENTIAL.
FT DISULFID	94	126	POTENTIAL.
FT DISULFID	160	180	POTENTIAL.
FT DISULFID	169	201	POTENTIAL.
FT DISULFID	233	252	POTENTIAL.
FT DISULFID	241	273	POTENTIAL.
FT DISULFID	304	323	POTENTIAL.
FT DISULFID	312	344	POTENTIAL.
FT DISULFID	378	397	POTENTIAL.
FT DISULFID	386	418	POTENTIAL.
FT DISULFID	443	462	POTENTIAL.
FT DISULFID	451	483	POTENTIAL.
FT DISULFID	507	527	POTENTIAL.
FT DISULFID	516	548	POTENTIAL.
FT DISULFID	592	612	POTENTIAL.
FT DISULFID	601	633	POTENTIAL.
FT DISULFID	675	694	BY SIMILARITY.
FT DISULFID	677	694	BY SIMILARITY.
FT DISULFID	696	705	BY SIMILARITY.
FT DISULFID	728	726	BY SIMILARITY.
FT DISULFID	709	741	BY SIMILARITY.
FT DISULFID	731	748	BY SIMILARITY.
FT DISULFID	750	759	BY SIMILARITY.
FT DISULFID	762	773	BY SIMILARITY.
FT DISULFID	810	830	POTENTIAL.
FT DISULFID	819	851	POTENTIAL.
FT DISULFID	1233	1244	BY SIMILARITY.
FT DISULFID	1238	1253	BY SIMILARITY.
FT DISULFID	1255	1264	BY SIMILARITY.
FT DISULFID	1450	1461	BY SIMILARITY.
FT DISULFID	1455	1471	BY SIMILARITY.
FT DISULFID	1473	1482	BY SIMILARITY.
FT DISULFID	1489	1500	BY SIMILARITY.
FT DISULFID	1494	1510	BY SIMILARITY.
FT DISULFID	1512	1521	BY SIMILARITY.
FT DISULFID	1718	1731	BY SIMILARITY.
FT DISULFID	1725	1740	BY SIMILARITY.
FT DISULFID	1742	1751	BY SIMILARITY.
FT CARBOHYD	390	390	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	659	659	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	764	764	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	814	814	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC	1648	1651	Missing (in isoform 3). /FtId=VSP_001369.
FT VARSPLIC	1783	1793	Missing (in isoform 2 and isoform 3). /FtId=VSP_001370.
FT CONFLICT	1129	1131	RTI -> SIL (IN REF. 1; AAA48586).
FT SEQUENCE	1955	AA; 211411	MW; B4DEB27C23422581 CRC64;

Query Match 20.8%; Score 796.5; DB 1; Length 1955;
 Best Local Similarity 30.6%; Pred. No. 1.3e-42;
 Matches 221; Conservative 99; Mismatches 271; Indels 131; Gaps 25;
 QY 2 IKITPRPDSADGMLLYNGKRVPGSPTNLANRQPDITFGLVGRPEFRFDAGSGMATIR 61
 Db 1294 IAMEFRATELSGLLYNGQ-----NRKDFISLALVGFVLELRENTSGSGVIT 1342
 QY 62 HPTPLALGHFTVTLRLSLTQGLSLLVGLDPLAVNGTSGKFGQGLDLNELVLYGYP-DYGA 120
 Db 1343 SKRVPEPKWHQLVVNRNRRSGLAV-DGEHFVSGESPTGDLNLDLFLVGGAPEDQWA 1401
 QY 121 I--PKAGLSSGFIGVNR-----ELRIQGBEIVFHDNLNLTAKHGISHC--PTCRDRFC 167
 Db 1402 VVAERTAAATVGLKGSIRLLDVANNQMYDLREKGSVLY-----GSGVGECDNDPCHNPFC 1455
 QY 168 QNGGQCHDSSESSYVCPAGFIGSRCEHSQALHCHPEACGPDPATCVNRPDGRGTYCRCH 227
 Db 1456 HFGASCHVYKAEAMFHCLELSHYTGTCA-DERNPCDPTPCHISATCLVLPEG-GAMCACP 1513

QY 2 IKITPPSADCMILYNGQKRVGSPNLANRQPDFISGLVGRGFRFFRFDAGSGMATIR 61
Db 673 ISMEFRASNDGLPLVQWTEKGGK-----LAFYRPS-----EGYVELRPHGCVMDVIT 721
QY 62 HPTPLALGHFTVILLRSITOGSLIVGLAPVNGTSQKFGQGLDNEELYVGGYPDYG-- 119
Db 722 SKTLIKPGNHHVYVGNRRRSGMLSDVGEPHLIGESPPTDGLNLDLDTLFLGGTPEDEMT 781
QY 120 -AIPKAGLSSGFIGCVRELRTOGEEIVPHDLN-----LTAHGISHCPT--CDRRCONG 170
Db 782 LVTERTTATKQLQGICIRLLDNN--LIYDLQERSNDVLYGSGVGGCGNPPCNPCPKR 838
QY 171 GQCHDSBSSVYVCVCPAGTGRSC--EHSOALHCHPEACGPDATCVNRPDGRGTYTCRHL 228
Db 839 GCHMKEAEMFHCSVGFSGPTCADKHNR--CDPNPCHOSANCVLPFG--GSKCECPM 894
QY 229 GRSGLRCE-----EGVYVTPISGAGSYLALPAL-----TWTHHELRLDVYFKPLADP 277
Db 895 GREGELSERVSAEODQOKAFIPFNGL--SYLENMNGIHTFVSDDLQKLSMEVIFLAKDN 953
QY 278 GYLLPFGKGSQVDFVSLAVMGHLEFRYELGSLAVRSAPLALGRWHRYSAERLNK 337
Db 954 GMIFYNGQKTDGRGDFVSLNLRDGLYLFKVDLKGAAVLRSKAPIPLNVVWVTVRNGR 1013
QY 338 DGLRNV-----GGRPVLRSPKSGGLMLHTLLLYLGGVPEPVPPLSPATWSAHRGCV 391
Db 1014 KGLMKINKDELVSGBSPKSRKAP--HTALNLRKEAFYVGGAPPDKNFARAAGIISGFTGAI 1071
QY 392 GEVYNGKRLDLYSFLGSGIQGOCYDSSPCRQFQHGATCMPAGEYEFQCLCRDGF 451
Db 1072 QKLSLK-----SIPLLKKNIRNAMEIS-----NFRW 1098
QY 452 DLCEHEENPCQLRPELGHGTC-----QGTRCLCLPFGSRRGCOGSGHGAESDHWLGG 507
Db 1099 HACTRTRPCQ-----NGGVCSPRLREYDCMCRGSPGQCEKA-----LE-- 1139
QY 508 GGNAPGOYGFYHDGFLAPFHVFRSLRPEVPEFTELEVRVSTASGLLLMGVGVGSA 567
Db 1140 -EKSASSEVAFNGRTEIYHNTVTRSEKAVQVNYFEMGIKTEATKGLIMS-----GKI 1194
QY 568 GQGKDFISGLQDHLVFRYQSGEARLYVSEDDINDGEWHRYTALRGRGSIQYDGE 627
Db 1195 AKSDIYALAVDGGVQMTYDLSGKVPVTRSTIPVNTQWVRKANKRHHGYGLQVGNRA 1254
QY 628 LVSGRSPGNVNAKGVYIGGAPDVAATLGGR-----FSSGITGCVKMLVLSHRAPGAP 683
Db 1255 PVTGSSPFAATQDLDGALMLGGIEKLA--PGNRLPKAYSTGTGCIKQVVIDR----- 1306
QY 684 PPQLDLQHRAGAGANTRPCPS 705
Db 1307 --OELQLVEDALNPTILHCPA 1326
RESULT 6
UN52 CAEBEL STANDARD; PRT; 3375 AA.
AC Q06561; 018261; 018263; Q9XTD2; Q9XIT5;
DT 01-JUN-1994 (Rel. 29, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Basement membrane proteoglycan precursor (Perlecan homolog)
DE (Uncoordinated protein 52).
GN UNC-52 OR ZC101.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Telodermata; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM A), AND FUNCTION.
RX MEDLINE=93339574; PubMed=8393416;
RA Rogalski T.M., Williams B.D., Mullen G.P., Moerman D.G.,
RT "Products of the unc-52 gene in Caenorhabditis elegans are homologous
to the core protein of the mammalian basement membrane heparan

RT sulfate proteoglycan.";
RL Genes Dev. 7:1471-1484 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Percy C.M., Baynes C.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RA Durbin R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Probable role in myofibrillar assembly and/or attachment
of the myofibrillar lattice to the cell membrane. May be an
extracellular anchor for integrin receptors in muscle.
CC -!- SUBCELLULAR LOCATION: Extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=4;
CC Name=a;
CC IsoId=Q06561-1; Sequence=Displayed;
CC Note=No experimental confirmation available;
CC Name=a;
CC IsoId=Q06561-2; Sequence=VSP_007195, VSP_007196;
CC Name=b;
CC IsoId=Q06561-3; Sequence=VSP_007191, VSP_007192;
CC Note=No experimental confirmation available;
CC Name=c;
CC IsoId=Q06561-4; Sequence=VSP_007193, VSP_007194, VSP_007195,
VSP_007196;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Found in the basement membrane of all
contractile tissues. It is concentrated over muscle dense bodies
and m-lines which are associated with beta-integrin.
CC -!- DEVELOPMENTAL STAGE: Synthesized early in embryogenesis.
CC -!- SIMILARITY: Contains 3 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 7 laminin EGF-like domains.
CC -!- SIMILARITY: Contains 3 laminin G-like domains.
CC -!- SIMILARITY: Contains 2 laminin IV domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
of the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L13458; AAA28156.1; -
DR EMBL; Z93375; CAB07567.1; -
DR EMBL; Z93395; CAB07567.1; JOINED.
DR EMBL; Z93375; CAB07568.1; -
DR EMBL; Z93395; CAB07568.1; JOINED.
DR EMBL; Z93395; CAB07569.1; -
DR EMBL; Z93395; CAB07569.1; JOINED.
DR EMBL; Z93395; CAB07704.1; -
DR EMBL; Z93395; CAB07706.1; -
DR EMBL; Z93375; CAB07706.1; JOINED.
DR EMBL; Z93395; CAB07707.1; -
DR EMBL; Z93375; CAB07707.1; JOINED.
DR EMBL; Z93395; CAB07708.1; -
DR EMBL; Z93375; CAB07708.1; JOINED.
DR HSSP; P01130; ILDR.
DR WormPep; ZC101.2a; CE15028.
DR WormPep; ZC101.2b; CE15030.
DR WormPep; ZC101.2c; CE15034.
DR WormPep; ZC101.2e; CE18424.
DR GO; GO:0005578; C:extracellular matrix; IEP.
DR GO; GO:0030239; P:myofibril assembly; IEP.
DR InterPro; IPR008985; Conserved_like_unc_52.
DR InterPro; IPR01881; EGF_Ca.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR007110; Ig-Like.

DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003598; Ig C2.
 DR InterPro; IPR00034; Laminin B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR Pfam; PF00047; Ig; 16.
 DR Pfam; PF00052; laminin_B; 2.
 DR Pfam; PF00053; laminin_EGF; 5.
 DR Pfam; PF00057; ldl_recept_a; 3.
 DR PRINTS; PD00261; LDLRECEPTOR.
 DR ProDom; PD003031; Laminin_B; 2.
 DR SMART; SM00181; EGF; 5.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SM00180; EGF_Lam; 6.
 DR SMART; SM00409; IG; 17.
 DR SMART; SM00408; IGG2; 17.
 DR SMART; SM00281; LamB; 1.
 DR SMART; SM00282; LamG; 3.
 DR SMART; SM00192; LDLa; 3.
 DR PROSITE; PS00022; EGF_1; 7.
 DR PROSITE; PS01185; EGF_2; 2.
 DR PROSITE; PS00026; EGF_3; 3.
 DR PROSITE; PS00835; IG_LIKE; 17.
 DR PROSITE; PS00025; LAM_G DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
 DR PROSITE; PS01209; LDLRA_1; 3.
 DR PROSITE; PS00068; LDLRA_2; 3.
 KW Proteoglycan; Glycoprotein; Signal; Immunoglobulin domain; Repeat;
 KW Basement membrane; Extracellular matrix; Alternative splicing;
 KW Laminin EGF-like domain.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 3375 BASEMENT MEMBRANE PROTEOGLYCAN.
 FT DOMAIN 45 130 IG-LIKE C2-TYPE 1.
 FT DOMAIN 148 184 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 189 229 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 232 269 LDL-RECEPTOR CLASS A 3.
 FT DOMAIN 271 355 IG-LIKE C2-TYPE 2.
 FT DOMAIN 384 431 LAMININ EGF-LIKE 1 (INCOMPLETE).
 FT DOMAIN 432 441 LAMININ EGF-LIKE 2 (N-TERMINAL).
 FT DOMAIN 442 633 LAMININ DOMAIN IV 1.
 FT DOMAIN 634 666 LAMININ EGF-LIKE 2 (C-TERMINAL).
 FT DOMAIN 674 720 LAMININ EGF-LIKE 3 (INCOMPLETE).
 FT DOMAIN 721 730 LAMININ EGF-LIKE 4 (N-TERMINAL).
 FT DOMAIN 731 921 LAMININ DOMAIN IV 2.
 FT DOMAIN 922 954 LAMININ EGF-LIKE 4 (C-TERMINAL).
 FT DOMAIN 955 1004 LAMININ EGF-LIKE 5.
 FT DOMAIN 1011 1060 LAMININ EGF-LIKE 6.
 FT DOMAIN 1061 1111 LAMININ EGF-LIKE 7.
 FT DOMAIN 1126 1222 IG-LIKE C2-TYPE 3.
 FT DOMAIN 1226 1311 IG-LIKE C2-TYPE 4.
 FT DOMAIN 1319 1401 IG-LIKE C2-TYPE 5.
 FT DOMAIN 1410 1499 IG-LIKE C2-TYPE 6.
 FT DOMAIN 1503 1585 IG-LIKE C2-TYPE 7.
 FT DOMAIN 1588 1680 IG-LIKE C2-TYPE 8.
 FT DOMAIN 1690 1785 IG-LIKE C2-TYPE 9.
 FT DOMAIN 1793 1878 IG-LIKE C2-TYPE 10.
 FT DOMAIN 1886 1970 IG-LIKE C2-TYPE 11.
 FT DOMAIN 1973 2069 IG-LIKE C2-TYPE 12.
 FT DOMAIN 2073 2163 IG-LIKE C2-TYPE 13.
 FT DOMAIN 2260 2343 IG-LIKE C2-TYPE 14.
 FT DOMAIN 2349 2435 IG-LIKE C2-TYPE 15.
 FT DOMAIN 2446 2530 IG-LIKE C2-TYPE 16.
 FT DOMAIN 2532 2713 IG-LIKE C2-TYPE 17.
 FT DOMAIN 2783 2960 LAMININ G-LIKE 1.
 FT DOMAIN 2961 3093 GLU-RICH.
 FT DOMAIN 2972 3066 THR-RICH.
 FT DOMAIN 3180 LAMININ G-LIKE 3.
 FT DISULFID 66 114 BY SIMILARITY.
 FT DISULFID 149 161 BY SIMILARITY.
 FT DISULFID 156 174 BY SIMILARITY.
 FT DISULFID 168 183 BY SIMILARITY.

FT DISULFID 190 202 BY SIMILARITY.
 FT DISULFID 197 215 BY SIMILARITY.
 FT DISULFID 209 224 BY SIMILARITY.
 FT DISULFID 233 246 BY SIMILARITY.
 FT DISULFID 240 259 BY SIMILARITY.
 FT DISULFID 253 288 BY SIMILARITY.
 FT DISULFID 285 964 BY SIMILARITY.
 FT DISULFID 957 971 BY SIMILARITY.
 FT DISULFID 974 983 BY SIMILARITY.
 FT DISULFID 986 1002 BY SIMILARITY.
 FT DISULFID 1011 1021 BY SIMILARITY.
 FT DISULFID 1013 1027 BY SIMILARITY.
 FT DISULFID 1030 1039 BY SIMILARITY.
 FT DISULFID 1042 1058 BY SIMILARITY.
 FT DISULFID 1061 1059 BY SIMILARITY.
 FT DISULFID 1063 1079 BY SIMILARITY.
 FT DISULFID 1082 1091 BY SIMILARITY.
 FT DISULFID 1094 1109 BY SIMILARITY.
 FT DISULFID 1152 1200 BY SIMILARITY.
 FT DISULFID 1338 1384 BY SIMILARITY.
 FT DISULFID 1435 1481 BY SIMILARITY.
 FT DISULFID 1527 1573 BY SIMILARITY.
 FT DISULFID 1618 1663 BY SIMILARITY.
 FT DISULFID 1719 1767 BY SIMILARITY.
 FT DISULFID 1814 1861 BY SIMILARITY.
 FT DISULFID 1907 1954 BY SIMILARITY.
 FT DISULFID 1998 2053 BY SIMILARITY.
 FT DISULFID 2099 2147 BY SIMILARITY.
 FT DISULFID 2195 2242 BY SIMILARITY.
 FT DISULFID 2284 2329 BY SIMILARITY.
 FT DISULFID 2374 2420 BY SIMILARITY.
 FT DISULFID 2467 2514 BY SIMILARITY.
 FT CARBOHYD 1422 1422 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2476 2476 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2950 2950 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 3143 3143 N-LINKED (GLCNAC. .) (POTENTIAL).
 Query Match 13.4%; Score 511; DB 1; Length 3375;
 Best Local Similarity 20.9%; Pred No. 2 6e-24;
 Matches 179; Conservative 140; Mismatches 332; Indels 204; Gaps 29;
 QY 1 EKTKFRPDSADGMLLYNQKRV-----GPTMLANRQPDFISFGLVGRPERFRPAGS 55
 Db 2557 DIBETLNTANPKGIIFET--KRINSGDLLATPYDTHHEAKITDYGTV---LYEFDIGN 2610
 QY 56 GMATRHPHTPLALGHFHTVLLRSLRSTOGLSIVGDLA-----PVNAGTSGKFGQGLDLN 107
 Db 2611 GRQIVETNPINPNEWNVKIKNDKQVVIQLNDESATIRQHTNPLPSLSTG-----VN 2664
 QY 108 BELYLGGYDPAIPKAGLSGFGICVRELRIOGEEIVFHDLNLTAKGISHCPTCRDRPC 167
 Db 2665 REVFVIGGRHE-----PTNEANDFRGIISQVVLGSHYVGLGDARIPSSVWVYDACAATNLC 2719
 QY 168 QNGGQCHDSES--SYVVCVPAGFTGSRCHESQALHCHPEACCPDATCVNRPDGRGYTCRC 226
 Db 2720 LAGNCANVNHGFSCECAEFHGEYQW--RNSCHDESCN-TGICLDNEE--SWQCVC 2775
 QY 227 HGRSGLRCEEGVTTTP-SLSGAGSYLALPALNTHHELRDLDFEKLAPDG---VLLF 282
 Db 2776 PLGTTGLRCEKTEIPQPLGFTSDTSLAVKRPVKE---SIRMKLRPOADSDSHILMYF 2832
 QY 283 SGGKSGPVDFVSLAMVGHLEFRYELGSLAVLSARPLALGRWHRVSAERLNKDGSLR 342
 Db 2833 ASDYGSNTKQYVSLSLIANQVV-----LTVRRPKVEQKIRSETLEAGEL-IDVAVR 2883
 QY 343 VNGRPFVLRSSPKSQGLNHLHTL-----LYLGGVPEVPLSPATNMAHRRGCVGEVSN 397
 Db 2884 QAGNALVMTVDGNQVSTIETDTLKEPTEIFIGGLPPGLN-SFDDVVEQSFQGCYIELLN 2942
 QY 398 GKRLDITVSLGSGQ-IGQCYS-----SPC 422
 Db 2943 SQDVL--QNLSSGDISSCESQPPVEEDDTTITTTTEPEAVIEPTTTEPTTTEPI 3000
 BY SIMILARITY.

FT DOMAIN 1506 1710 LAMININ G-LIKE 1.
 FT DOMAIN 1713 1749 EGF-LIKE 4, CALCIUM-BINDING.
 FT DOMAIN 1753 1935 LAMININ G-LIKE 2.
 FT DOMAIN 1937 1972 EGF-LIKE 5, CALCIUM-BINDING.
 FT DOMAIN 1973 2011 EGF-LIKE 6, CALCIUM-BINDING.
 FT DOMAIN 2012 2044 EGF-LIKE 7, CALCIUM-BINDING.
 FT DOMAIN 2046 2081 EGF-LIKE 8, CALCIUM-BINDING.
 FT DOMAIN 2087 2120 LAMININ EGF-LIKE.
 FT DOMAIN 2475 2527 GFS.
 FT DISULFID 1370 1381 BY SIMILARITY.
 FT DISULFID 1375 1412 BY SIMILARITY.
 FT DISULFID 1414 1423 BY SIMILARITY.
 FT DISULFID 1430 1441 BY SIMILARITY.
 FT DISULFID 1435 1450 BY SIMILARITY.
 FT DISULFID 1452 1461 BY SIMILARITY.
 FT DISULFID 1470 1481 BY SIMILARITY.
 FT DISULFID 1475 1491 BY SIMILARITY.
 FT DISULFID 1493 1504 BY SIMILARITY.
 FT DISULFID 1717 1728 BY SIMILARITY.
 FT DISULFID 1722 1737 BY SIMILARITY.
 FT DISULFID 1739 1748 BY SIMILARITY.
 FT DISULFID 1941 1952 BY SIMILARITY.
 FT DISULFID 1946 1961 BY SIMILARITY.
 FT DISULFID 1963 1972 BY SIMILARITY.
 FT DISULFID 1976 1987 BY SIMILARITY.
 FT DISULFID 1981 1999 BY SIMILARITY.
 FT DISULFID 2010 2010 BY SIMILARITY.
 FT DISULFID 2018 2031 BY SIMILARITY.
 FT DISULFID 2033 2043 BY SIMILARITY.
 FT DISULFID 2050 2065 BY SIMILARITY.
 FT DISULFID 2052 2068 BY SIMILARITY.
 FT DISULFID 2070 2080 BY SIMILARITY.
 FT MOD RES 1954 1954 HYDROXYLATION (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 838 838 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1173 1173 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1213 1213 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1308 1308 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1318 1318 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1640 1640 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1704 1704 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1761 1761 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2044 2044 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2173 2173 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2192 2192 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2382 2382 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2472 2472 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2504 2504 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3313 AA; 359348 MW; B11DA09517288764 CRC64;
 Query Match 11.5%; Score 441; DB 1; Length 3313;
 Best Local Similarity 23.9%; Pred. No. 6.8e-20;
 Matches 187; Conservative 87; Mismatches 256; Indels 252; Gaps 42;
 QY 61 RHPPLALGHE-HVTLRLRSITQSSLI-----VGD-----LAPVNGTSQK- 100
 DB 1277 RELSPL-LGHLEGVAAVLATFEDVFNIONQTDVGGTVLNVSVFALAP-RGAGAA 1334
 QY 101 ---FOGLDNEBELLYLGGYDYGAIKAGLSS-----GFTGCVREL 137
 DB 1335 GPFWSSEELQEQLY-----VRAALAAARLSDLVLPDDNVCLREPCENYMKCVSVL 1385
 QY 138 R-----LQGEIIVPHDNLNTAKGISHCIP-----TCRDRPQNGGQCHDSE 177
 DB 1386 RFDSSAPFLASATLFRPQIOPFAGLRCKRCPGFGTDFCETELDLCYSPCKNGACARR 1445
 QY 178 SSSYVCVCPAGFTGRCE-HSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRS--GLR 234
 DB 1446 -GGYTCVCRPRFTGEDCELDTEAGRCVPCVCRNGGTCNAFNG-GFRCQCQFAGAFGRPR 1503
 QY 235 CSEGVVITTPSLSGSYLALPALTWHEHLELDVRYKPLAPDGVLLFESGKSGPVEDFV 294
 DB 1504 CEVAARSFPFS-----SFWNFRGLQRHF-LTSLSLSFATVQPSG-LLFYNGRLNKEKDFL 1556

QY 295 SLAVYGGHLEFRYELGSSGLAVLRSAP--LALGRWHEVSAERLNKDGSLRVNGRPLVRS 352
 DB 1557 ALELVAGQVRLTYTSGESSTVVSFVPGGLSDGQWHTVHLRYNKPRTDALGGAQ----- 1611
 QY 353 SPGK-----SGLNLHHTLLYLGVEPVS 375
 DB 1612 GPKDKVAVLSVDDCNVAALRFGAEIGNYSCAAAGVQTSKSKSLDLTGPFLLLGGV--PNL 1670
 QY 376 PLS-PATNMSARFRGCVGEVSVNGKRLDLTYSFILGSGIQCYDSSPCERQPCQHGATCM 434
 DB 1671 PENFVSRKD--FICWEDLHIDGRVDM-AFVANNGT-----TAGCQ 1711
 QY 435 PAGYEFQCLCRDGFKBLCEHENPCQLRBPCLHGTCQ-----GTRCLCLPQFSGRPCQ 490
 DB 1712 AKSHF-----CASG-----PKKNGGLCSERMGFGFCDCPFVGFGGKDCR 1749
 QY 491 QGSGHGIAESDWHLEGGC-----GNDAPGOYGAFFHDDGLFAPPGHVFSRSLPEVPEFI 544
 DB 1750 LTMAL-----PVHFGNGTLSWDFGNDMP-----VSVFWYL 1780
 QY 545 ELEVTSTASGLLLWQGYEVG-----EAGQKDFISLGLDGHLYFRYQLSGEARLV 597
 DB 1781 GLSPFRATKGLM--QVQLGPHSVLLCKLDQGLLSVTLRSRASHAVH-----LLL 1829
 QY 598 SEDPINDGEWH--RVTALRE--GRRG-----SIQVDGELVSGRSPGENVAVNAKGSYIIG 649
 DB 1830 DQMTVSDGRWEDLRLELQEEPGRGRHHIFMVSLDFTLFDQTMWAMGSELEGLKVKLHVQ 1889
 QY 650 GAPDVATVITGGFSSGIVGCVKLVLSHARPGA---PPQPLDLOHR--AAQGAN--TRPC 703
 DB 1890 GPPSSKEEG---PQGLVGCIGQVWVTGTFPFGSSALPPP-----SHRINVEPCGCTVTPC 1941
 QY 704 PS 705
 DB 1942 AS 1943
 RESULT 8
 NX2A_RAT STANDARD; PRT; 1715 AA.
 ID NX2A_RAT
 AC Q63374; Q63375;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Neurexin 2-alpha precursor (Neurexin II-alpha).
 GN NRXN2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., VARIANT LEU-434, AND ALTERNATIVE SPLICING.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=92320296; PubMed=1621094;
 RA Ushkaryov Y.A., Petrenko A.G., Geppert M., Suedhof T.C.;
 RT "Neurexins: synaptic cell surface proteins related to the alpha-
 RT latrotoxin receptor and laminin";
 RL Science 257:50-56(1992).
 RN [2]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Brain;
 RX MEDLINE=95209856; PubMed=7695896;
 RA Ullrich B., Ushkaryov Y.A., Suedhof T.C.;
 RT "Cartography of neurexins: more than 100 isoforms generated by
 RT alternative splicing and expressed in distinct subsets of neurons.";
 RL Neuron 14:497-507(1995).
 RN [3]
 RP INTERACTION WITH NEUREXOPHILIN 1.
 RX MEDLINE=99074239; PubMed=9856994;
 RA Missler M., Hammer R.E., Suedhof T.C.;
 RT "Neurexophilin binding to alpha-neurexins. A single LNS domain
 RT functions as an independently folding ligand-binding unit.";

J. Biol. Chem. 273:34716-34723(1998).
 [4]
 INTERACTION WITH ALPHA-DYSTROGLYCAN.
 MEDLINE=21363578; PubMed=11470830;
 Sugita S, Saito F, Tang J, Satz J, Campbell K., Suedhof T.C.;
 "A stoichiometric complex of neuexins and dystroglycan in brain.";
 J. Cell Biol. 154:435-445(2001).
 -1- FUNCTION: Neuronal cell surface protein that may be involved in cell recognition and cell adhesion. May mediate intracellular signaling.
 -1- SUBUNIT: The laminin G-like domain 1 binds to neuexophilin 1. Isoforms alpha 2C bind to alpha-dystroglycan.
 -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 -1- ALTERNATIVE PRODUCTS:
 -1- Event=Alternative promoter;
 Comment=A number of isoforms, alpha-type (shown here) and beta-type, are produced by use of alternative promoters.
 Beta-type isoforms differ from alpha-type isoforms in their N-terminus;
 Event=Alternative splicing; Named isoforms=1;
 Comment=At least 216 isoforms may be produced by alternative splicing. There is a combination of five alternatively spliced extracellular domains at sites 1 to 5, each consisting of modular sequences (A-C) that seem to be used independently. Additional isoforms may derive from a minor cytoplasmic splice site 6. Beta isoforms (AC Q63376) share the combination of alternatively spliced domains at sites 4, 5 and 6;
 Name=Alpha1A2A3A4A5A;
 IsoId=Q63374-1; Sequence=Displayed;
 -1- TISSUE SPECIFICITY: Brain (neuronal synapse).
 -1- SIMILARITY: Contains 6 laminin G-like domains.
 -1- SIMILARITY: Contains 3 EGF-like domains.
 -1- SIMILARITY: Belongs to the neuexin family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMEL; M96376; AAA41706.1; -;
 DR EMEL; M96376; AAA41707.1; -;
 DR PIR; C40228; C40228.
 DR HSSP; Q63373; 1C4R.
 DR InterPro; IPR008985; Cona like lec_gl.
 DR InterPro; IPR006209; EGF Like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin G.
 DR InterPro; IPR003585; Neuexin-Like.
 DR Pfam; PF00008; EGF; 3.
 DR Pfam; PF00054; laminin_G; 5.
 DR SMART; SM00294; 4.1m; 1.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00282; LamG; 6.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; FALSE_NEG.
 DR PROSITE; PS50026; EGF_3; 3.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 6.
 DR Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
 KW Polypeptide; Alternative splicing; Alternative promoter usage;
 KW Polyorphism.
 FT SIGNAL 1 29
 FT CHAIN 30 1715
 FT DOMAIN 30 1639
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1640 1660
 FT DOMAIN 1661 1715
 FT CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 206 206
 FT LAMININ G-LIKE 1.
 FT DOMAIN 202 242
 FT EGF-LIKE 1.
 FT DOMAIN 289 486
 FT LAMININ G-LIKE 2.
 FT DOMAIN 493 686
 FT LAMININ G-LIKE 3.
 FT DOMAIN 690 727
 FT EGF-LIKE 2.

FT	DOMAIN	732	907	LAMININ G-LIKE 4.
FT	DOMAIN	921	1096	LAMININ G-LIKE 5.
FT	DOMAIN	1099	1136	EGF-LIKE 3.
FT	DOMAIN	1140	1348	LAMININ G-LIKE 6.
FT	DOMAIN	1371	1374	POLY-THR.
FT	DOMAIN	1377	1380	POLY-THR.
FT	DOMAIN	1447	1450	POLY-PRO.
FT	DOMAIN	1647	1650	POLY-ALA.
FT	CARBOHYD	60	60	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	295	295	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	338	338	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	844	844	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1239	1239	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	250	259	Missing (in some isoforms Alpha 1B). /FTid=VSP_003509.
FT	VARSPLIC	250	283	Missing (in some isoforms Alpha 1C). /FTid=VSP_003510.
FT	VARSPLIC	385	399	Missing (in some isoforms Alpha 2C). /FTid=VSP_003511.
FT	VARSPLIC	393	399	Missing (in some isoforms Alpha 2B). /FTid=VSP_003512.
FT	VARSPLIC	797	806	DCLRVGCAPS -> G (in some isoforms Alpha 3C). /FTid=VSP_003513.
FT	VARSPLIC	807	809	Missing (in some isoforms Alpha 3B). /FTid=VSP_003514.
FT	VARSPLIC	1256	1285	Missing (in some isoforms Alpha 4B). /FTid=VSP_003515.
FT	VARSPLIC	1421	1614	Missing (in some isoforms Alpha 5B). /FTid=VSP_003516.
FT	VARSPLIC	1667	1715	DEGSYVDQSRNYSNSAOSNGAVVKEKAPAPKTPSKAKK NKDEYVY -> CRKSPREKLLPSSAOGLDLAKACVC RCRAICIAKLEBERGGGREGGRCMQIYIKNK (in some isoforms Alpha 6). /FTid=VSP_003517. N -> L.
FT	VARIANT	434	434	
FT	SEQUENCE	1715	AA; 59	59BFB18661F3DB15 CRC64;
Query Match			11.0%;	Score 421; DB 1; Length 1715;
Best Local Similarity			21.6%;	Fred. No. 5.9e-19;
Matches 193;	Conservative 114;	Mismatches 274;	Indels 312;	Gaps 37;
QY	2	IKITRPSADQMLLYN-GQKRVPSPTNLANRQPDFISFGLVGGRRFRFPDAGSGMATI	60	
Db	517	ISLDFRTPEPNGLLFSQRRRAGAGVGHSSSRADYFAMELLDGLYLLLDMDGSGGKIL	576	
QY	61	RHPT-PLALGHFHTLLRSLTQGLIVGDIA-FVNGTISQKFGGLDNEELYLGSPDY	118	
Db	577	RASRRKVDNGEYVDFQDRGRKSGISVNSRSTPPLAT--GESEVIDLESELYLGLPEG	634	
QY	119	GAIP-----KAGLSSGFCVRELRIOGBEIVFHDLNLTAGHSHCP-----TC	162	
Db	635	GRVDLPLPEVWTAALRAGYVGVVDRDLFDGRSRDLRGLAEAGGAVGVAPFCRSRETLKQC	694	
QY	163	RDRPQNGGCHDSESSSVVCV-C-PAGFTGSCFE-----HSCA	199	
Db	695	ASAPCRNGGICREG-WNRFVCDICIGTFLGRYCEREATVLSYDGGMYKIMLPNAMHTPA	753	
QY	200	-----VNRP-----LHCHEPCGGDATC-----	213	
Db	754	EDVSLRFMSQRAYGLMMATTSRESADTLRLDGGQMKLVNLDCLRVCAPSAAKGPE	813	
QY	214	-----VNRP-----	217	
Db	814	TLFAGHKLNDNEWHTLRRVRRGKSLQLSVDNVTVEGQWAGATRLRLEFFHNIETGIMTERRF	873	
QY	218	-----DGRVYTCRHLG-----RSLGRCEGVTVVTPFSLGAGS	251	
Db	874	ISVPSNFGHLSGLVFNQGFMDQCKDGDITYCELNARFGLRAIVADPVTFKRSR---S	930	
QY	252	YLALPALWTHRELRLDVEFKPLADPVLGLFSGKSGPVEDFVSLAMVGGHLEFRVELGS	311	
Db	931	YLALATL-QAYASMLFFQFKTAPDGLLLFNSGNG---NDFIVIELVKGYHYVFDLGN	986	

312 GLAVLR--SAEPLALGRWHRVSAER-LNKDGLSLRVNGRVPVLRSSFGKSGOGLNHLTLLYL 368
 987 SPSLMKGNBPKVNDQWVSVRSRPGNVNHLKID-SRTVTQHSNG-ARNLDLKGLYLI 1044
 369 GGYEVSPLSPATNMAH--FRGCVGEVSNKRLDLYSLGSGOGLGQCY-----DSS 420
 1045 GGLSKNM-FSNLEKLVASRDGFGCLASVDLNGRPLDLIADAL--HRIGQVVERGCDGPEST 1101
 421 PCRQPCQHGATCMPAGEYEFQCLCRDGFKGBDLCEHSENPQOLRPECLHGFGCOGTRCLC 480
 1102 TCTEESCANQVCLQW-----DGF-----TCDCT----- 1126
 481 LPFSGPRCOQ-----GSHGIIAESSDWHLESGSGNDAPQGYAFVDDGFLAFPGHVF 533
 1127 MTSYGGPVCNDPFTTIFKGGALIIYH-----PPNDRPSTR-----MDRLAVGSETHQR 1177
 534 SRSLPEVFEFIEVFTSTASGILLMQGVEVGEAGQKDFISLQDGHVFRYOLGSGE 593
 1178 SAVL-----VRVDSASGL-----GDYLQLHIDQGTGVIFNVGFTDD 1213
 594 ARLVSEDP--INDGHRVHTALRGRGSIQVD--GEELVSGRSPGPNVAV----- 640
 1214 --ITIDEPNAIVSDGXHVVRFTFSGNATLQVDSWPVNERYPAGNFENRERLAIARQIP 1271
 641 -----NAKGSVYIGGAPDVATLTGGRFSSGITGCVN 672
 1272 YRLGRVVDEWLDKGRQLTIFNSQAAIKIGRRDQ-----GRPFQGVSLYIN 1319

RESULT 9

NX2A HUMAN STANDARD; PRT; 1712 AA.
 AC Q9P2S2; Q9Y2D6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Neurexin 2-alpha precursor (Neurexin II-alpha).
 GN NXN2 OR KIAA0921.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Fetal brain;
 RA Seki N., Yoshikawa T., Azuma T., Muramatsu M., Saito T.;
 RT "Human neurexin II.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Rowen L., Madan A., Qin S., Baradarani L., Birditt B., Bloom S.,
 RA Burke J., Dors M., Fleetwood P., Kaur A., Madan A., Neebitt R.,
 RA Pate D., Hood L.;
 RT "Sequencing of human neurexin II gene.";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2), AND REVISIONS.
 RC TISSUE=Brain;
 RX MEDLINE=22158633; PubMed=12168954;
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara C., Nagase T.;
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual
 duration of 330 KIAA cDNA clones.";
 RL DNA Res. 9:99-106(2002).
 RN [4]
 RP SEQUENCE OF 272-1712 FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=9246063; PubMed=10231032;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hiroseawa M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara C.;
 RT "Prediction of the coding sequences of unidentified human genes. XIII.
 The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro.";

RL DNA Res. 6:163-70(1999).
 CC -1- FUNCTION: Neuronal cell surface protein that may be involved in
 CC cell recognition and cell adhesion. May mediate intracellular
 CC signaling.
 CC -1- SUBUNIT: The laminin G-like domain 1 binds to neurexophilin 1.
 CC Specific isoforms bind to alpha-dystroglycan (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative promoter;
 CC Comment=A number of isoforms, alpha-type (shown here) and
 CC beta-type (AC P58401), are produced by use of alternative
 CC promoters. Beta-type isoforms differ from alpha-type isoforms
 CC in their N-terminus;
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q9P2S2-1; Sequence=Displayed;
 CC Name=2; Synonyms=Alpha-2B;
 CC IsoId=Q9P2S2-2; Sequence=VSP_003505, VSP_003506, VSP_003507,
 CC VSP_003508;
 CC -1- TISSUE SPECIFICITY: Predominantly expressed in brain.
 CC -1- SIMILARITY: Contains 6 laminin G-like domains.
 CC -1- SIMILARITY: Contains 3 EGF-like domains.
 CC -1- SIMILARITY: Belongs to the neurexin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB035266; BAA94075.1; -;
 DR EMBL; AC044790; AAK69154.1; -;
 DR EMBL; AB023138; BAA76765.2; ALT_INIT.
 DR HSSP; O63373; 1C4R.
 DR Genew; HGNC:8009; NRXN2.
 DR MIM; 600566; -;
 DR InterPro; IPR008985; Cona_like Lec-gl.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin G.
 DR InterPro; IPR003585; Neurexin-like.
 DR Pfam; PF00008; EGF_3.
 DR Pfam; PF00054; Laminin_G; 5.
 DR SMART; SM00294; 4.lm; 1.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00282; LamG; 6.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; FALSE_NEG.
 DR PROSITE; PS50026; EGF_3; 3.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 6.
 DR Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
 KW Glycoprotein; Alternative splicing; Alternative promoter usage;
 FT SIGNAL 1 28
 FT CHAIN 29 1712 NEUREXIN 2-ALPHA.
 FT DOMAIN 29 1636 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1637 1657 POTENTIAL.
 FT DOMAIN 1658 1712 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 29 206 LAMININ G-LIKE 1.
 FT DOMAIN 202 242 EGF-LIKE 1.
 FT DOMAIN 289 486 LAMININ G-LIKE 2.
 FT DOMAIN 493 686 LAMININ G-LIKE 3.
 FT DOMAIN 690 727 EGF-LIKE 2.
 FT DOMAIN 732 904 LAMININ G-LIKE 4.
 FT DOMAIN 918 1033 LAMININ G-LIKE 5.
 FT DOMAIN 1096 1133 EGF-LIKE 3.
 FT DOMAIN 1137 1345 LAMININ G-LIKE 6.
 FT DOMAIN 1368 1371 POLY-THR.
 FT DOMAIN 1374 1377 POLY-THR.
 FT DOMAIN 1444 1447 POLY-PRO.
 FT DOMAIN 1644 1647 POLY-ALA.

FT CARBOHYD 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT FT 338 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 841 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT FT 1236 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 260 Missing (in isoform 2).
 FT /FTid=VSP_003505.
 FT VARSPLIC 393 Missing (in isoform 2).
 FT /FTid=VSP_003506.
 FT VARSPLIC 797 DCLRVCAPF -> G (in isoform 2).
 FT /FTid=VSP_003507.
 FT VARSPLIC 1253 Missing (in isoform 2).
 FT /FTid=VSP_003508.
 FT SEQUENCE 1712 AA; 184980 MW; E488EC404DAD10B0 CRC64;
 Query Match 11.0%; Score 420.5; DB 1; Length 1712;
 Best Local Similarity 21.6%; Pred. No. 6.3e-19;
 Matches 192; Conservative 114; Mismatches 275; Indels 309; Gaps 37;

QY 2 IKITFRPDSADGMLLYN-GOKRVPGSPTNLANRQPDFISGLVGRPEFRFDAGSGMARI 60
 Db 517 ISLDFRITTEFNGLLLFSQRRAGGAGSHSAGRADIYFAMELLDGHLYLLLDWGGGKYL 576
 QY 61 RHPT-PLALGHFHTVLLRSLTQGLSVGLDIA-PVNGTSGKQFQGLDNEELVLYGYPDY 118
 Db 577 RASSRKVNDGEWCHVDPRDGRKGSISVNSRSTFFLAT--GDSEILDSELYLGLGPEP 634
 QY 119 GAIP-----KAGLSSGFGICVRELRIGBEIVFHDNLNLTAGH:SHCP-----TC 162
 Db 635 GRVDLPPEVPEVTAALRAGYVGVVDRDFIDGRSDLRGLAQAQVAVFAPSRETLKQC 694
 QY 163 RDRPQMGQCHDSESSYVCV-PAGFTGSRCE-----HSQA 199
 Db 695 ASAPCRNGGVCREG-WNRFCDCIGTGTFLGRVCERATVLSYDGSVMYKIMLPNAMHTEA 753
 QY 200 -----LHCHPEACGPD----- 210
 Db 754 EDVSLRFMSQRAYGLMATTRESADTLRELDGGQMKLTVNLDCLRVGCAPSKEGPELTF 813
 QY 211 -----ARCVNRP----- 217
 Db 814 AGHKLNDNEHWTVRVRGKSLQSLVDNVTVEGQAGAHMRLEBFHNIEITGIMTERRFISV 873
 QY 218 -----DGRYTCRHLG-----RSLGRCGEVTVTTPSLGAGSYLA 254
 Db 874 VPSNFIHLSGLVFNQFVMDQCKDGDITVYCELMARFLRAIVADPVTFKRS----SYLA 930
 QY 255 LPALTNTTHELRDVEKPLAPDGVLLFSGKSGPVEDFVSLAMVGGHLEFRYELGSLGA 314
 Db 931 LATL-QAYASHELFPQKTAAPDGLLLFNSGNG--NDFIVIELVKYIHYVFDLNGPS 986
 QY 315 VLR--SAPPLALGRHVRSAER-LNKDGLRVNGRPPVLRSSPKGSLMLHLLYLGCV 371
 Db 987 LMKGNSDKPVNDQWNVVSRDPGNVHTLKID-SRTVVTQHSNG-ARNLDDKGLYITGGL 1044
 QY 372 EPSVPLSPATWMSAH---FRGCVGEVSNVGRKLDLTYSLGSGIGQCY-----DSBPCE 423
 Db 1045 SKNM-FSNLPLKVAARDGFCQCLASVDLNGRLPLDIADAL--HRIGQVRCGDGPTTCT 1101
 QY 424 RQPCOHGATCPAGEYEFQCLCRDGFKDLCEHEENPCQLREPCCLHGGTQGTGTRCLCLPG 483
 Db 1102 EESCANQGVCLQQW-----DGF-----TCDCI-----MTS 1126
 QY 484 FSGPRCQQ-----GSGHGIAESDWHLEGGGNDAPQGYAFHDDGFLAFPGHVSRS 536
 Db 1127 YGFPVNDPGTYTFPKGKALITYTW-----PPNDRPSTR-----MDRLVGFSTHQRSAV 1177
 QY 537 LPVPEETIELEVRTSTAGLLWQVGEVGEAGQKDFISLGLQDGHVFRYQLGSGEARL 596
 Db 1178 L-----VRVDSASGL-----GDYLLQHLIDQGTGVIFNVGTTD--I 1211
 QY 597 VSEDP---INDGEHRVTAALREGRGSLQVD-----GELVSGRFPGNVAV----- 640
 Db 1212 TIDEPNAIVSDGKYHVVRFTFRSGGNATLQVDSWPNRYFAGNFNDEKLAIRARQIPYRL 1271

QY 641 -----NAKGSVYICGAPDVAITLGTGRFSSGTTGCVKN 672
 Db 1272 GRVVDVLLDKGRQLTIFNSOAIKIGRDO-----GRPFQGVSGLYYN 1316

RESULT 10
 NX1A_HUMAN
 ID NX1A_HUMAN STANDARD; PRT; 1477 AA.
 AC Q9ULB1; O60323; Q9C079; Q9C080; Q9C081; Q9H3M2; Q9UDM6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Neurexin 1-alpha precursor (Neurexin I-alpha).
 GN NRXN1 OR KIAA0578
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Seki N., Yoshikawa T., Azuma T., Saito T., Muramatsu M.;
 RT "Human neurexin I-alpha."
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 105-1477 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98290545; PubMed=9628581;
 RA Nagura T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura T., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro."
 RL DNA Res. 5:31-39(1998).
 RN [3]
 RP SEQUENCE OF 441-1121 FROM N.A.
 RA Mead K., Wohlmann P., Minges B.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 1083-1477 FROM N.A.
 RA Rowen L., Madan A., Qin S., Baradarani L., Birditt B., Bloom S.,
 RA Dors M., Dickhoff R., Fleetwood P., Harrison G., Kaur A., Madan A.,
 RA Nesbitt R., Traicoff R., Hood L.;
 RT "Sequencing of human chromosome 2 neurexin I gene."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Neuronal cell surface protein that may be involved in
 CC cell recognition and cell adhesion. May mediate intracellular
 CC signaling.
 CC -!- SUBUNIT: The cytoplasmic C-terminal region binds to CASK. Laminin
 CC G-like domain 2 binds to NXP1. Specific isoforms bind to alpha-
 CC dystroglycan and to alpha-latroxin. Binds syt11 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative promoter;
 CC Comment-A number of isoforms are produced by use of alternative
 CC promoters. The alpha and beta (AC P58400) isoforms differ in
 CC their N-terminus;
 CC Event-Alternative splicing; Named isoforms=1;
 CC Comment-A number of isoforms may be produced;
 CC Name=1;
 CC IsoId=Q9ULB1-1; Sequence=Displayed;
 CC -!- TISSUE SPECIFICITY: Heart and brain.
 CC -!- PTM: N- and O-glycosylated (By similarity).
 CC -!- SIMILARITY: Contains 6 laminin G-like domains.
 CC -!- SIMILARITY: Contains 3 EGF-like domains.
 CC -!- SIMILARITY: Belongs to the neurexin family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

CC EMBL; AB035356; BAA87821.1; -
CC EMBL; AB011150; BAA25504.1; -
CC EMBL; AC007452; AAF03536.1; -
CC EMBL; AC068725; AAG59602.1; -
CC EMBL; AC069550; AAG38120.1; -
CC EMBL; AC078994; AAK06387.1; -
CC EMBL; AC068715; AAG59642.1; -
CC HSP; O63373; 1C4R.
CC Genew; HGNC:8008; NRXN1.
CC MIM; 600545; -
CC InterPro; IPR000152; Asx hydroxyl S.
DR InterPro; IPR008985; ConAlike lec_g1.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001791; Laminin G.
DR InterPro; IPR003585; Neurexin-like.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00054; laminin_G; 6.
DR SMART; SM00294; 4.1m; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00282; LamG; 6.
DR PROSITE; PS00010; ASX HYDROXYL; 1.
DR PROSITE; PS00022; EGF 1; FALSE_NEG.
DR PROSITE; PS01186; EGF 2; FALSE_NEG.
DR PROSITE; PS50026; EGF 3; 3.
DR PROSITE; PS50025; Lam_G_DOMAIN; 6.
KW Signal; transmembrane; Repeat; Cell adhesion; EGF-like domain;
KW Glycoprotein; Alternative splicing; Alternative promoter usage.
FT SIGNAL 1 30
FT CHAIN 31 1477
FT DOMAIN 31 1401
FT DOMAIN 1402 1422
FT TRANSMEM 1423 1477
FT DOMAIN 1423 1477
FT DOMAIN 31 217
FT DOMAIN 213 256
FT DOMAIN 283 473
FT DOMAIN 480 572
FT DOMAIN 676 713
FT DOMAIN 718 891
FT DOMAIN 905 1080
FT DOMAIN 1083 1120
FT DOMAIN 1126 1294
FT DOMAIN 1324 1327
FT DOMAIN 1409 1412
FT POLY-THR. 1412
FT CARBOHYD 125 125
FT CARBOHYD 190 190
FT CARBOHYD 790 790
FT CARBOHYD 1223 1223
SQ SEQUENCE 1477 AA; 161882 MW; F7845FB428B1A683 CRC64;

Query Match 10.8%; Score 412.5; DB 1; Length 1477;
Best Local Similarity 20.8%; Pred. No. 1.7e-18;
Matches 200; Conservative 113; Mismatches 287; Indels 361; Gaps 41;

Qy 2 IKITRFRDSDMLY-NGQKVPYSPWLANRPPDFISFGLVGGRRPFRFAGSGMATI 60
Db 504 ISFDFTTEPNGLIFSHGKPRQKAKHPQMKVDFPAEMLDGHLVLLLEMGSGTIKI 563

Qy 61 RHPTPLAL-----GHFFHTVTLRLSLTQGSLLYVGDLL-APVNGTSQKFGGLDLNELYLG 113
Db 564 K-----ALLAKVNDGEWYHVDQDGRSGTISVNTLRTPY--TAPGESEILDDBELYLG 616

Qy 114 GYDYGAIKPKAG-----LSSFGICVTELRIGQEEIVFH--DLNLTAFHGISHC 159
Db 617 GLPE-----NKAGLVFPEVWVALLNYGVICRDLFDIGGSKDIRQMAEVQSTAGKPKSC 672

Qy 160 -----PTCRDRPCNGGQCHDSESSSYVCCP--AGFTGSRCE----- 195
Db 673 SKETAKCLSNPKKNGMCRDG-WNRVYVCCSGTGYLGRSCEREATVLSYDGSFMFKIQI 731

Qy 196 ---HSQA-----LHCHEPAC-- 207
Db 732 FVWHTAEADVSLRFRQRAYGILMATTSRDSADTLRLLELDAGRKVLKTNLDCIRINCNS 791

Qy 208 ---GPDATCV--NRPDGRGYTCR----- 225
Db 792 SKGPETLFAGYNLNDNWHVTRVRRRGSKLKLTVDVDDQAMTGMAGDTRLEFHNIEFGI 851

Qy 226 -----CHL-GRSGLR--CEEGVTVTT 243
Db 852 ITERRYLSSVPSNFCHLQSLTFNGMAYIDLCKNGDIDYCELNARFGFRNIADPVTFTK 911

Qy 244 PLSGAGSYLALPALTNTHHEURLRVEFKPLAPDGVLLFSGKSGPVEDFVSLAMVGGHL 303
Db 912 KS-----SYVALATL-QAYTSMHLPFOKTTSLDGLILYN---SGDNDFIIVVELVKG 962

Qy 304 EFRYELGSLAVLR--SABPLALGRWHRVSAER-----LNKDGSLRVNNGRPFVLR 352
Db 963 HVVFDLNGANLTKGSSNKPLNDQWNVMSIDTISNLHTVKIDTKITQITAG----- 1016

Qy 353 SPKSGQLMLHTLLYLGTV--EPSVPLSPATNMSAHRFCVGEVSKRLDLYSFLGS 410
Db 1017 ----ARNLDKSDLYIGGVAKETYSKLPKLVHAKGFGQCLASVDLNGRLPDL----- 1065

Qy 411 QGIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKDLCEHEENPCQLRPFCLHG 470
Db 1066 ----ISDALFCN-----GQIE-----RG--CEGPTTCQ--EDSCSNQ 1095

Qy 471 GTC-----QGTRECLC-LPGFSGPRCQO-----GSHGIAESDWHLLEGGGNDAPQCYGA 518
Db 1096 GVCLQOWDFSCDMSMTSFGPLCNDPGTTYIFSKGGQITYKW-----PPNDRPSTRA- 1149

Qy 519 YFDDQFLAPPGHVFRSLEPEVETIELR---VRTSTASGLLLWQGVGEAGQGGKDFIS 575
Db 1150 ---DRLAIGF-----STVQKEAVLVRVDSGSL-----GDYLE 1179

Qy 576 LGLQDGHVFRYQLGSGEARLVSDEP--INDGEHRVTVALLREGRGSIQVDGELVSRGP 634
Db 1180 LHIHQKIGVKENVGTDIAIESNAIINDKHYVVRFRSGGNATLOVDSWPFVIERPA 1239

Qy 635 GPVAV-NAKGSVYIGG----- 650
Db 1240 GRQITFNQATIIIGKEQGFQFQGLSGLYNGLKVMAENDANAIIVGNVRLGE 1299

Qy 651 -----APDVATLTGRRFFSGGITGVKLVLSHARPAPPPPLDLOHRAQAGANRPPC 704
Db 1300 VPSSMTESATAMQSEMSTSIMETTTLATSTARPKPTKEPISTQTTDDILVASAEC 1359

Qy 705 S 705
Db 1360 S 1360

RESULT 11
NKXIA_BOVIN STANDARD; PRT; 1530 AA.
ID NKXIA_BOVIN
AC Q28146;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neurexin 1-alpha precursor (Neurexin I-alpha).
GN NRXN1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN (1)
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain.
RX MEDLINE=95209856; PubMed=7698986;
RA Ullrich B., Ushkaryov Y.A., Suedhof T.C.;
RT "Cartography of neurexins: more than 1000 isoforms generated by

RT alternative splicing and expressed in distinct subsets of neurons."
RL Neuron 14:497-507(1995).
RN [2]
RP INTERACTION WITH NEUREXOPHILIN.
RX MEDLINE=96285495; PubMed=8699246;
RA Petrenko A.G., Ullrich B., Missler M., Krasnoperov V., Rosahl T.W.,
RA Suedhof T.C.;
RT "Structure and evolution of neuropsychophyllin."
RL J. Neurosci. 16:4360-4369(1996).
CC -1- FUNCTION: Neuronal cell surface protein that may be involved in
CC cell recognition and cell adhesion. May mediate intracellular
CC signaling.
CC -1- SUBUNIT: The cytoplasmic C-terminal region binds to CASK. Binds
CC SYTL1. Laminin G-like domain 2 binds to NXP1. Isoforms 9 and
CC alpha-4c bind to alpha-dystroglycan. Isoform alpha-4c binds to
CC alpha-latroxin (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative promoter;
CC Comment=A number of isoforms, alpha-type (shown here) and
CC beta-type, are produced by use of alternative promoters.
CC Beta-type isoforms differ from alpha-type isoforms in their
CC N-terminus;
CC Event=Alternative splicing; Named isoforms=12;
CC Comment=additional isoforms seem to exist. There is probably
CC more than 96 isoforms. There is a combination of five
CC alternatively spliced domains at sites 1 to 5, each consisting
CC of modular sequences (A-G) that seem to be used independently.
CC For splice site 1 additional splice modules might be possible.
CC thus increasing the number of possible isoforms. Beta isoforms
CC (AC Q28142) share the combination of two alternatively spliced
CC domains at sites 4 and 5. Experimental confirmation may be
CC lacking for some isoforms.
CC Name=1; Synonyms=Alpha-1A2A3A4A5A;
CC IsoId=Q28146-1; Sequence=displayed;
CC Name=2; Synonyms=Alpha-1B;
CC IsoId=Q28146-2; Sequence=VSP_003478;
CC Name=3; Synonyms=Alpha-1C;
CC IsoId=Q28146-3; Sequence=VSP_003474, VSP_003478;
CC Name=4; Synonyms=Alpha-1D;
CC IsoId=Q28146-4; Sequence=VSP_003477;
CC Name=5; Synonyms=Alpha-1E;
CC IsoId=Q28146-5; Sequence=VSP_003474, VSP_003477;
CC Name=6; Synonyms=Alpha-1F;
CC IsoId=Q28146-6; Sequence=VSP_003476, VSP_003478;
CC Name=7; Synonyms=Alpha-1G;
CC IsoId=Q28146-7; Sequence=VSP_003475;
CC Name=8; Synonyms=Alpha-2B;
CC IsoId=Q28146-8; Sequence=VSP_003480;
CC Name=9; Synonyms=Alpha-2C;
CC IsoId=Q28146-9; Sequence=VSP_003479;
CC Name=10; Synonyms=Alpha-3B;
CC IsoId=Q28146-10; Sequence=VSP_003481;
CC Name=11; Synonyms=Alpha-4B;
CC IsoId=Q28146-11; Sequence=VSP_003482;
CC Name=12; Synonyms=Alpha-5B;
CC IsoId=Q28146-12; Sequence=VSP_003483;
CC -1- PTM: N- and O-glycosylated (By similarity).
CC -1- SIMILARITY: Contains 6 laminin G-like domains.
CC -1- SIMILARITY: Contains 3 EGF-like domains.
CC -1- SIMILARITY: Belongs to the neuexin family.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; LI4855; AA474123.1;
CC PIR; I45944; I45944.
CC HSSP; Q63373; 1C4R.

DR	InterPro; IPR000152; Asx_hydroxyl_S.
DR	InterPro; IPR008985; ConA_like_1ec_gl.
DR	InterPro; IPR006209; EGF_like.
DR	InterPro; IPR006210; IEGF.
DR	InterPro; IPR001791; Laminin G.
DR	InterPro; IPR003585; Neuexin-like.
DR	Pfam; PF00008; EGF_2.
DR	Pfam; PF00054; laminin_G; 6.
DR	SMART; SM00294; 4.lm; 1.
DR	SMART; SM00181; EGF; 2.
DR	SMART; SM00282; LamG; 6.
DR	PROSITE; PS00010; ASX_HYDROXYL; 1.
DR	PROSITE; PS00022; EGF_1; FALSE_NEG.
DR	PROSITE; PS01186; EGF_2; FALSE_NEG.
DR	PROSITE; PS00026; EGF_3; 3.
DR	PROSITE; PS00025; LamG_DOMAIN; 6.
KW	Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
KW	Glycoprotein; Alternative splicing; Alternative promoter usage.
FT	SIGNAL 1 30
FT	CHAIN 31 1530
FT	NEUREXIN 1-ALPHA.
FT	DOMAIN 31 1454
FT	EXTRACELLULAR (POTENTIAL).
FT	POTENTIAL.
FT	TRANSMEM 1455 1475
FT	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 1476 1530
FT	LAMININ G-LIKE 1.
FT	DOMAIN 31 212
FT	EGF-LIKE 1.
FT	DOMAIN 213 256
FT	LAMININ G-LIKE 2.
FT	DOMAIN 299 496
FT	LAMININ G-LIKE 3.
FT	DOMAIN 503 695
FT	EGF-LIKE 2.
FT	DOMAIN 699 736
FT	LAMININ G-LIKE 4.
FT	DOMAIN 741 914
FT	LAMININ G-LIKE 5.
FT	DOMAIN 928 1103
FT	EGF-LIKE 3.
FT	DOMAIN 1106 1143
FT	LAMININ G-LIKE 6.
FT	DOMAIN 1149 1347
FT	POLY-THR.
FT	DOMAIN 1377 1380
FT	DOMAIN 1462 1465
FT	CARBOHYD 125 125
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 190 190
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 813 813
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 1246 1246
FT	VARSPPLIC 258 258
FT	Missing (in isoform 3 and isoform 5).
FT	FTID=VSP_003474.
FT	VARSPPLIC 258 293
FT	Missing (in isoform 7).
FT	FTID=VSP_003475.
FT	VARSPPLIC 264 273
FT	Missing (in isoform 6).
FT	FTID=VSP_003476.
FT	VARSPPLIC 274 293
FT	Missing (in isoform 4 and isoform 5).
FT	FTID=VSP_003477.
FT	VARSPPLIC 278 293
FT	Missing (in isoform 2, isoform 3 and
FT	isoform 6).
FT	FTID=VSP_003478.
FT	VARSPPLIC 395 409
FT	Missing (in isoform 9).
FT	FTID=VSP_003479.
FT	VARSPPLIC 403 409
FT	Missing (in isoform 8).
FT	FTID=VSP_003480.
FT	VARSPPLIC 806 815
FT	DCIRINCNS_ -> G (in isoform 10).
FT	FTID=VSP_003481.
FT	VARSPPLIC 1263 1292
FT	Missing (in isoform 11).
FT	FTID=VSP_003482.
FT	VARSPPLIC 1426 1428
FT	Missing (in isoform 12).
FT	FTID=VSP_003483.
FT	SEQUENCE 1530 AA; 167937 MW; 8A4E4A75C4EC03D1 CRC64;
EQ	Query Match 10.6%; Score 405.5; DB 1; Length 1530;
EQ	Best Local Similarity 20.5%; Pred. No. 4.9e-18;
EQ	Matches 203; Conservative 112; Mismatches 285; Indels 391; Gaps 42;
QY	2 IKITRPDSADGMILY-NGQKVPGPSPTLANRQPDFTSFGLVGRPFPRFDAGSGMATI 60
DB	527 ISPDFRTTEPHEGHLIFSHGKPRHXDKAKPKMKVDFFAIEMLDGHLVLLDDMGSGTIKI 586
.QY	61 RHPTPLAL-----GHFTVTLLRSLTQGSLLVGLDLPVNVGTSQKFGLDINRELYLG 113
DB	587 K-----ALQKYNQDGEWHVDFQDRSGTISVNTLRTPY---TAFGESILDDELXYLG 639

CC -1- SIMILARITY: Contains 3 EGF-like domains.
 CC -1- SIMILARITY: Belongs to the neuexin family.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch)

CC EMBL; M96374; AAA41704.1; -;
 DR PIR; A40228; A40228.
 DR HSP; O63373; 1C4R.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR008985; Con1_like_lec_gl.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin G.
 DR InterPro; IPR003585; Neuexin-like.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00054; Laminin G; 6.
 DR SMART; SM00294; 4.lm; 1.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00282; LamG; 5.
 DR PROSITE; PS0010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; FALSE_NEG.
 DR PROSITE; PS50026; EGF_3; 3.
 DR PROSITE; PS50025; LAM_G DOMAIN; 6.
 KW Signal; transmembrane; Repeat; Cell adhesion; EGF-like domain;
 KW Glycoprotein; Alternative splicing; Alternative promoter usage.
 FT SIGNAL 1 30
 FT CHAIN 31 1514 NEUREXIN 1-ALPHA.
 FT DOMAIN 31 1438 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1439 1459 POTENTIAL.
 FT DOMAIN 1460 1514 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 31 217 LAMININ G-LIKE 1.
 FT DOMAIN 213 255 EGF-LIKE 1.
 FT DOMAIN 283 480 LAMININ G-LIKE 2.
 FT DOMAIN 487 679 LAMININ G-LIKE 3.
 FT DOMAIN 683 720 EGF-LIKE 2.
 FT DOMAIN 725 898 LAMININ G-LIKE 4.
 FT DOMAIN 912 1087 LAMININ G-LIKE 5.
 FT DOMAIN 1090 1127 EGF-LIKE 3.
 FT DOMAIN 1133 1331 LAMININ G-LIKE 6.
 FT DOMAIN 1361 1364 POLY-THR.
 FT DOMAIN 1446 1449 POLY-ALA.
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 797 797 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1230 1230 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPPLIC 258 258 Missing (in some isoforms Alpha 1C).
 FT VARSPPLIC 258 258 /FTid-VSP_003486.
 FT VARSPPLIC 258 277 Missing (in some isoforms Alpha 1G).
 FT VARSPPLIC 274 277 /FTid-VSP_003487.
 FT VARSPPLIC 379 393 Missing (in some isoforms Alpha 1D).
 FT VARSPPLIC 387 393 /FTid-VSP_003488.
 FT VARSPPLIC 387 393 Missing (in some isoforms Alpha 2C).
 FT VARSPPLIC 387 393 /FTid-VSP_003489.
 FT VARSPPLIC 790 799 Missing (in some isoforms Alpha 2B).
 FT VARSPPLIC 790 799 /FTid-VSP_003490.
 FT VARSPPLIC 790 799 DCIRINCNS -> G (in some isoforms Alpha 3B).
 FT VARSPPLIC 790 799 /FTid-VSP_003491.
 FT VARSPPLIC 1247 1276 Missing (in some isoforms Alpha 4B).
 FT VARSPPLIC 1410 1411 /FTid-VSP_003492.
 FT VARSPPLIC 1410 1411 Missing (in some isoforms Alpha 5B).
 FT VARSPPLIC 1410 1411 /FTid-VSP_003493.
 FT SEQUENCE 1514 AA; 166167 MW; D99492202968DE14 CRC64;

Query Match 10.5%; Score 403.5; DB 1; Length 1514;
 Best Local Similarity 20.4%; Pred. No. 6.5e-18;

Qy	Db	Mismatches	Conservative	Indels	Gaps
2	IKITFRSDSADGMLLY-NQOKRVP	113	285	391	42
511	ISPDFTTEPNGLLFSHGRPRQ				
61	RHPTPLAL-----GHFHTV				
571	K-----ALQKVNDGEWYHVD				
114	GYPDYGAIPKAG-----L				
624	GLPE-----NKAGLVFPEV				
160	-----PTCRDRPCQNGCQ				
680	SRETAKPCLSNPKCKNGMCRD				
196	-----HSQA-----				
739	PVMVHTEAEVDSLRFSORAY				
208	--GPDATCV--NRPDGRGYTC				
799	SKGPETLPAGYNLNDNEWH				
226	-----CHL-CESGLR--				
859	ITERRYLSSVPSNFIGHLQ				
244	PSLSGAGSYLALPALTNH				
919	KS-----SYVALATL-QA				
304	EFRYELSGLAIVR--SAPL				
970	HYVFLGNGANLIGSSNKP				
353	SPKSGQLNHLTLLVGGV--				
1024	-----ARNLDKSDLIYG				
411	QGIGCYDSSPCERQPCQ				
1073	-----ISDALFCN-----				
471	GTC-----OGTRCLC-L				
1103	GVCLQQWDGFCDCSMTS				
519	YFHDDGFLAFPGHVF				
1157	---DRLAIGF-----				
576	LGLQDGLHVFYQLGSGE				
1187	LHIHQKVGKFNKFN				
631	GRSPGNVAV-----				
1247	GNDNERLAIAQRIPYR				
651	-----				
1307	LYYNGLKLVAENDAN				
675	LHSARFPAPPQPLDLQ				
1367	TSTARRGKPTKEPISQ				

RESULT 13
 CLR3_MOUSE

CLR3 MOUSE STANDARD; PRT; 3301 AA.
 09A1Z10_QPESDQ;
 28-FEB-2003 (Rel. 41, Created)
 28-FEB-2003 (Rel. 41, Last sequence update)
 10-OCT-2003 (Rel. 42, Last annotation update)
 Cadherin EGF LAG seven-pass G-type receptor 3 precursor.
 CELSR3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGES.
 RC STRAIN=C57BL/6;
 RX MEDLINE=21839555; PubMed=11850187;
 RA Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;
 RT "Developmental expression profiles of Celsr (Flamingo) genes in the
 mouse.";
 RL Mech. Dev. 112:157-160(2002).
 RN [2]
 RP SEQUENCE OF 2574-3046 FROM N.A., AND DEVELOPMENTAL STAGE.
 RX MEDLINE=21534980; PubMed=11677057;
 RA Formstone C.J., Little P.F.R.;
 RT "Chromosomal localization of Celsr2 and Celsr3 in the mouse; Celsr3 is
 a candidate for the tippy (tip) lethal mutant on chromosome 9.";
 RL Mamm. Genome 11:392-394(2000).
 CC [3]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=20253755; PubMed=10790539;
 RA Formstone C.J., Barclay J., Rees M., Little P.F.R.;
 RT "Chromosomal localization of Celsr2 and Celsr3 in the mouse; Celsr3 is
 a candidate for the tippy (tip) lethal mutant on chromosome 9.";
 RL Mamm. Genome 11:392-394(2000).
 CC [4]
 RP FUNCTION: Receptor that may have an important role in cell/cell
 signaling during nervous system formation.
 CC [5]
 RP SUBCELLULAR LOCATION: Integral membrane protein.
 CC [6]
 RP TISSUE SPECIFICITY: Expressed in the CNS and in the eye.
 CC [7]
 RP DEVELOPMENTAL STAGE: Predominantly expressed in the CNS, the
 emerging dorsal root ganglia and cranial ganglia. In the CNS,
 expression is uniform along the rostrocaudal axis. No expression
 is detected until somite stages. Between E10 and E12, expression
 is strong in the marginal zone (MZ), and lower in the ventricular
 zone (VZ). At E15, expression is restricted to the brain and
 olfactory epithelium. In the brain, it is low in VZ but strong in
 external fields, particularly those with ongoing migration, such
 as the telencephalic cortical plate, the olfactory bulb, the
 cerebellum and the tectum. In the newborn and postnatal stages,
 expression is high in differentiated neuronal fields.
 CC [8]
 RP SIMILARITY: Belongs to family 2 of G-protein coupled
 receptors.
 CC [9]
 RP SIMILARITY: Contains 9 cadherin domains.
 CC [10]
 RP SIMILARITY: Contains 2 EGF-like domains.
 CC [11]
 RP SIMILARITY: Contains 1 laminin G-like domain.
 CC [12]
 RP SIMILARITY: Contains 1 GPS domain.
 CC [13]
 RP CAUTION: Ref.2 sequence differs from that shown due to frameshifts
 in positions 2575 and 2578.
 CC [14]
 RP This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC [15]
 RP EMBL; AF427498; AAL25099.1;
 DR EMBL; AF188752; AAG17057.1; ALT_FRAME.
 DR MED; MG1.1858236; Celsr3.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000126; Cadherin.
 DR InterPro; IPR008985; ConA_like_1ec_gl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR006209; EGF_like.

InterPro; IPR000832; GPCR_secretin.
 InterPro; IPR001879; hormn_receptor.
 InterPro; IPR006210; IEFG.
 InterPro; IPR002049; Laminin_EGF.
 InterPro; IPR001791; Laminin_G.
 InterPro; IPR000203; PKD_cys_rich.
 DR Pfam; PF00002; 7tm2; 1.
 DR Pfam; PF00028; cadherin; 9.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF01825; GPS; 1.
 DR Pfam; PF02793; HRM; 1.
 DR Pfam; PF00053; laminin_EGF; 1.
 DR Pfam; PF00054; laminin_G; 2.
 DR PRINTS; PR00205; CADHERIN.
 DR PRINTS; PR00249; GPCRSECRETIN.
 DR SMART; SM00112; CA; 9.
 DR SMART; SM00181; EGF; 6.
 DR SMART; SM00303; GPS; 1.
 DR SMART; SM00008; Hormr; 1.
 DR SMART; SM00282; LamG; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00232; CADHERIN_1; 7.
 DR PROSITE; PS00268; CADHERIN_2; 8.
 DR PROSITE; PS00022; EGF_1; 5.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS00026; EGF_3; 6.
 DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
 DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
 DR PROSITE; PS00227; G_PROTEIN_RECEP_F2_3; 1.
 DR PROSITE; PS00251; G_PROTEIN_RECEP_F2_4; 1.
 DR PROSITE; PS00221; GPS; 1.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 2.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
 KW Developmental protein; Hydroxylation; Signal.
 FT SIGNAL 1 31 POTENTIAL.
 FT CHAIN 32 3301 CADHERIN EGF LAG SEVEN-PASS G-TYPE
 RECEPTOR 3.
 FT DOMAIN 32 2531 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 2532 2552 1 (POTENTIAL).
 FT DOMAIN 2553 2563 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 2564 2584 2 (POTENTIAL).
 FT DOMAIN 2585 2592 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 2593 2613 3 (POTENTIAL).
 FT DOMAIN 2614 2634 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 2635 2655 4 (POTENTIAL).
 FT DOMAIN 2656 2673 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 2674 2694 5 (POTENTIAL).
 FT DOMAIN 2695 2715 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 2717 2737 6 (POTENTIAL).
 FT DOMAIN 2738 2744 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 2745 2765 7 (POTENTIAL).
 FT DOMAIN 2766 3301 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 317 424 CADHERIN 1.
 FT DOMAIN 425 536 CADHERIN 2.
 FT DOMAIN 537 642 CADHERIN 3.
 FT DOMAIN 643 747 CADHERIN 4.
 FT DOMAIN 748 849 CADHERIN 5.
 FT DOMAIN 850 952 CADHERIN 6.
 FT DOMAIN 953 1058 CADHERIN 7.
 FT DOMAIN 1059 1160 CADHERIN 8.
 FT DOMAIN 1161 1257 CADHERIN 9.
 FT DOMAIN 1366 1424 EGF-LIKE 1.
 FT DOMAIN 1426 1460 EGF-LIKE 2.
 FT DOMAIN 1464 1503 EGF-LIKE 3.
 FT DOMAIN 1504 1708 LAMININ G-LIKE 1.
 FT DOMAIN 1711 1747 EGF-LIKE 4.
 FT DOMAIN 1751 1933 LAMININ G-LIKE 2.
 FT DOMAIN 1935 1971 EGF-LIKE 5.
 FT DOMAIN 1972 2002 EGF-LIKE 6.
 FT DOMAIN 2003 2042 EGF-LIKE 7.
 FT DOMAIN 2044 EGF-LIKE 8.

2085 2118 LAMININ EGF-LIKE.
 2468 2520 GPS.
 2720 2724 POLY-LEU.
 1370 1381 BY SIMILARITY.
 1375 1412 BY SIMILARITY.
 1414 1423 BY SIMILARITY.
 1430 1441 BY SIMILARITY.
 1435 1450 BY SIMILARITY.
 1452 1459 BY SIMILARITY.
 1468 1479 BY SIMILARITY.
 1473 1489 BY SIMILARITY.
 1491 1502 BY SIMILARITY.
 1715 1726 BY SIMILARITY.
 1720 1735 BY SIMILARITY.
 1737 1746 BY SIMILARITY.
 1939 1950 BY SIMILARITY.
 1944 1959 BY SIMILARITY.
 1961 1970 BY SIMILARITY.
 1974 1985 BY SIMILARITY.
 1979 1997 BY SIMILARITY.
 1999 2008 BY SIMILARITY.
 2016 2029 BY SIMILARITY.
 2031 2041 BY SIMILARITY.
 2048 2063 BY SIMILARITY.
 2058 2066 BY SIMILARITY.
 2068 2078 BY SIMILARITY.
 1952 1952 HYDROXYLATION (POTENTIAL).
 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 838 N-LINKED (GLCNAC. . .) (POTENTIAL).
 1173 N-LINKED (GLCNAC. . .) (POTENTIAL).
 1213 N-LINKED (GLCNAC. . .) (POTENTIAL).
 1308 N-LINKED (GLCNAC. . .) (POTENTIAL).
 1318 N-LINKED (GLCNAC. . .) (POTENTIAL).
 1638 N-LINKED (GLCNAC. . .) (POTENTIAL).
 1702 N-LINKED (GLCNAC. . .) (POTENTIAL).
 1759 N-LINKED (GLCNAC. . .) (POTENTIAL).
 2042 N-LINKED (GLCNAC. . .) (POTENTIAL).
 2166 N-LINKED (GLCNAC. . .) (POTENTIAL).
 2185 N-LINKED (GLCNAC. . .) (POTENTIAL).
 2375 N-LINKED (GLCNAC. . .) (POTENTIAL).
 2465 N-LINKED (GLCNAC. . .) (POTENTIAL).
 2497 N-LINKED (GLCNAC. . .) (POTENTIAL).
 2713 L -> LR (IN REF. 2).
 3024 R -> P (IN REF. 2).
 3301 AA; 358455 MW; A6B18F2D7F4DEB6 CRC64;
 10.5%; Score 403; DB 1; Length 3301;
 23.0%; Pred. No. 1.7e-17;
 Matches 179; Conservative 91; Mismatches 263; Indels 244; Gaps 40;
 61 RHPTPLALGFH--HTVTLRLSLTQGSLLI-----VGD-----LAPVNGTSGGK- 100
 1277 RFLSPL-LGHELEGVAALAPPTEDVETFINQNDTVGGTVLVNVSFALAP-RGAGAGAA 1334
 101 ---FQGLDNLNELYGGYPDGAIPKAGLAS-----VRRALAAARSLLDVLPDDNVCLREPCENYMKCVSVL 137
 1335 GPWFSEELQELQY-----VRRALAAARSLLDVLPDDNVCLREPCENYMKCVSVL 1385
 138 R-----IQEEIVFDHNLNITAHGISHCP-----TCDRPCQNGGQCHDSE 177
 1386 RFDSSAPFLASTLFRIOIPIAGLRCRCPFGTDFCETELDLCSNPNCRNGGACARRE 1445
 178 SSSIVCVCPAGFTSRCE-HSQALHCHPEACGPATCVNRPDGRGYTCRCHLGRS--GLR 234
 1446 -GGYTCVCRPRFTD--CELDTEAGRCVPGVCRNGGTCINAPNG-GFRCCQCPAGAFEGPR 1501
 235 CEEGVTVVTPSLGAGSYLALPALTNTHELRDVEFKPLAPDGVLLFSGGKSPVEDFV 294
 1502 CEVAARSFPSP-----SFVFRGLRQRPH-LTSLSFATVQPSG-LLFYNGRLNEKHDFL 1554
 295 SLAMVGGHLEFRYELGSLAVRSNEP--LALGRHHRVSAERLNKDGSLRVNGRPFVLR 352
 1555 ALELVAGVRLTYTSGENTVTVPTVPGGLSDGQMHVTHLRYNKRPTDALGGAQ----- 1609

353 SPGK-----SQGLNHLTLLYLGCVPSV 375
 1610 GFSKDKVAVLSVDDCNVAVALQFGAEIGNYSCAAAGVQTSKKSLEDLTGLPLLLGGV-PNL 1668
 376 PLS-PATMSAHFRGCVGEVSVNGKRLDLTVSFLGSGQIGQCYSVSPCERQPCQHGATCM 434
 1669 PENFPVSHKD--FVGCWRDLHIDGRMDMA-AFYANNGT-----MAGCQ 1709
 435 PAGEYEFQCLCRDGFKGDGLCEHENPCOLREPCPLHGGTCQ-----GTRCLCLPFGSGPRCQ 490
 1710 AKSHF-----CASG-----PCXNNGFCSERWGGFCDCPVGFGKDCR 1747
 491 QSGHGIAESDWHLESGS-----GNDAPGGYGAFFHDDGFLAPFGHVFSRSLRSEVPEPI 544
 1748 LTMAR-----PYHQMGKTLSDWDFND-----MAVSVPEWYL 1778
 545 ELEVRTSPAGLLMQGVEVG-----BAGQKDFISLGLQDGHVFRYQLGSGEARLV 597
 1779 GLSFRTRATKGIIM--QVQLGPHSVLLCKLDRGLLSVTLNARASGHTVH-----LLL 1827
 598 SEDPTNGEWH--RVTALRE--GRGG---SIOVDGEBELVSRSPGPNVAVNAKGSVVYIG 649
 1828 DQMTVSDGRWHDLKLLELQEEPEGRRGHHIFVMSLDFTLFQDTMAMGGELQGLKVKQLHVG 1887
 650 GAPDVATLTGGRFSSGITGVKNLVLHRSARFGAPPQPPQLDLOHRAQAGAN-TRPCPS 705
 1888 GLPSSKEEG--HQGLVGCIOGVWIGTPTFGSSALLPPSHRVNVEPGCTVTFPCAS 1941
 RESULT 14
 ID SLT2 HUMAN STANDARD; PRT; 1529 AA.
 AC O94813; O95710; O95907;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Slit homologs 2 protein precursor (h-Slit-2).
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 OX [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
 RC TISSUE=Fetal lung;
 RX MEDLINE=99033071; PubMed=981312;
 RA Itoh A, Miyabayashi T, Ohno M, Sakano S;
 RT "Cloning and expressions of three mammalian homologues of Drosophila
 RT slit suggest possible roles for slit in the formation and maintenance
 RT of the nervous system".
 RL Brain Res. Mol. Brain Res. 62:175-186(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 3), AND TISSUE SPECIFICITY.
 RC TISSUE=Fetal brain, and Fetal kidney;
 RX MEDLINE=99279238; PubMed=10349621;
 RA Holmes G.P., Negus K., Burridge L., Raman S., Algar E., Yamada T.,
 RA Little M.H.;
 RT "Distinct but overlapping expression patterns of two vertebrate slit
 RT homologs implies functional roles in CNS development and
 RT organogenesis".
 RL Mech. Dev. 79:57-72(1998).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2), SEQUENCE OF 1122-1129, FUNCTION,
 RP SUBUNIT AND SUBCELLULAR LOCATION.
 RC TISSUE=fetal brain;
 RX MEDLINE=99200391; PubMed=10102268;
 RA Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,
 RA Tessier-Lavigne M., Kidd T.;
 RT "Slit proteins bind Robo receptors and have an evolutionarily
 RT conserved role in repulsive axon guidance".
 RL Cell 96:795-806(1999).
 CC -!- FUNCTION: Plays a role in axon guidance as a repulsive ligand for

Robo receptors preventing inappropriate midline crossing.
 -!- SUBUNIT: Binds robo proteins with high affinity.
 -!- SUBCELLULAR LOCATION: Secreted. The C-terminal cleavage protein is more diffusible than the larger N-terminal protein that is more tightly cell associated.
 -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=3;
 Name=1;
 IsoId=O94813-1; Sequence=Displayed;
 Name=2;
 IsoId=O94813-2; Sequence=VSP_050035, VSP_050036;
 Name=3;
 IsoId=O94813-3; Sequence=VSP_050036;
 -!- TISSUE SPECIFICITY: Fetal lung and kidney, and adult spinal cord.
 Weak expression in adult adrenal gland, thyroid, trachea and other tissues examined.
 -!- SIMILARITY: Contains 7 EGF-like domains.
 -!- SIMILARITY: Contains 23 leucine-rich (LRR) repeats.
 -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 -!- SIMILARITY: Contains 1 laminin G-like domain.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 DR EMEL; AB017168; BAA35185.1; -!
 DR EMEL; AFU55585; AAD04309.1; -!
 DR EMEL; AF133270; AAD25539.1; -!
 DR HSSP; P00743; ICCF.
 DR GEM; HGNC:11086; SLIT2.
 DR GEM; G03746; -!
 DR GO; GO:0005576; C:extracellular; IMP.
 DR GO; GO:0045499; F:chemorepellant activity; IMP.
 DR GO; GO:0005515; F:protein binding; TAS.
 DR GO; GO:0008347; F:glia cell migration; IMP.
 DR GO; GO:0075097; F:mesoderm migration; IMP.
 DR GO; GO:0080445; F:motor axon guidance; TAS.
 DR GO; GO:0007399; F:neurogenesis; TAS.
 DR GO; GO:0008038; F:neuronal cell recognition; TAS.
 DR GO; GO:0007608; P:olfaction; TAS.
 DR InterPro; IPR00152; Asx_hydroxyl_S.
 DR InterPro; IPR008985; Cys_knot_C.
 DR InterPro; IPR006207; EGF_2.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR003645; FOLN.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR000483; LRR_Cterm.
 DR InterPro; IPR000372; LRR_Nterm.
 DR InterPro; IPR003591; LRR_Typ.
 DR Pfam; PF00008; EGF_8.
 DR Pfam; PF00054; laminin_G; 1.
 DR Pfam; PF00560; LRR; 18.
 DR Pfam; PF01463; LRRCT; 4.
 DR Pfam; PF01462; LRRNT; 4.
 DR SMART; SM00041; CT_1.
 DR SMART; SM00179; EGF_Ca; 2.
 DR SMART; SM00274; FOLN; 2.
 DR SMART; SM00282; Lang; 1.
 DR SMART; SM00369; LRR_Typ; 8.
 DR SMART; SM00082; LRRCT; 4.
 DR SMART; SM00013; LRRNT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 2.
 DR PROSITE; PS01185; CTCK_1; 1.
 DR PROSITE; PS01225; CTCK_2; 1.
 DR PROSITE; PS00022; EGF_1; 9.
 DR PROSITE; PS01186; EGF_2; 7.

DR PROSITE; PS50026; EGF_3; 9.
 DR PROSITE; PS01187; EGF_CA; 2.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 1.
 KW Neurogenesis; Glycoprotein; Signal; Alternative splicing;
 KW EGF-like domain; Repeat; Leucine-rich repeat.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 1529 SLIT PROTEIN N-PRODUCT.
 FT CHAIN 31 1121 SLIT PROTEIN N-PRODUCT.
 FT CHAIN 1122 1529 SLIT PROTEIN C-PRODUCT.
 FT SITE 1121 1192 CLEAVAGE.
 FT REPEAT 54 777 LRR 1.
 FT REPEAT 78 101 LRR 2.
 FT REPEAT 102 125 LRR 3.
 FT REPEAT 126 149 LRR 4.
 FT REPEAT 151 173 LRR 5.
 FT REPEAT 174 197 LRR 6.
 FT REPEAT 199 226 LRR 7.
 FT REPEAT 299 322 LRR 8.
 FT REPEAT 323 346 LRR 9.
 FT REPEAT 347 370 LRR 10.
 FT REPEAT 372 394 LRR 11.
 FT REPEAT 395 418 LRR 12.
 FT REPEAT 511 531 LRR 13.
 FT REPEAT 532 556 LRR 14.
 FT REPEAT 557 580 LRR 15.
 FT REPEAT 582 604 LRR 16.
 FT REPEAT 605 628 LRR 17.
 FT REPEAT 629 652 LRR 18.
 FT REPEAT 753 776 LRR 19.
 FT REPEAT 777 799 LRR 20.
 FT REPEAT 800 823 LRR 21.
 FT REPEAT 824 847 LRR 22.
 FT REPEAT 849 872 LRR 23.
 FT DOMAIN 918 955 EGF-LIKE 1.
 FT DOMAIN 957 996 EGF-LIKE 2.
 FT DOMAIN 998 1034 EGF-LIKE 3.
 FT DOMAIN 1036 1074 (POTENTIAL).
 FT DOMAIN 1076 1112 EGF-LIKE 4.
 FT DOMAIN 1121 1157 EGF-LIKE 5.
 FT DOMAIN 1160 1333 EGF-LIKE 6.
 FT DOMAIN 1332 1368 LAMININ G-LIKE.
 FT DOMAIN 1453 1528 EGF-LIKE 7.
 FT DISULFID 922 933 CTCK.
 FT DISULFID 943 943 BY SIMILARITY.
 FT DISULFID 945 954 BY SIMILARITY.
 FT DISULFID 961 972 BY SIMILARITY.
 FT DISULFID 965 984 BY SIMILARITY.
 FT DISULFID 986 995 BY SIMILARITY.
 FT DISULFID 1002 1013 BY SIMILARITY.
 FT DISULFID 1007 1022 BY SIMILARITY.
 FT DISULFID 1024 1033 BY SIMILARITY.
 FT DISULFID 1040 1053 BY SIMILARITY.
 FT DISULFID 1047 1062 BY SIMILARITY.
 FT DISULFID 1064 1073 BY SIMILARITY.
 FT DISULFID 1080 1091 BY SIMILARITY.
 FT DISULFID 1085 1100 BY SIMILARITY.
 FT DISULFID 1102 1111 BY SIMILARITY.
 FT DISULFID 1125 1136 BY SIMILARITY.
 FT DISULFID 1130 1145 BY SIMILARITY.
 FT DISULFID 1147 1156 BY SIMILARITY.
 FT DISULFID 1336 1346 BY SIMILARITY.
 FT DISULFID 1341 1356 BY SIMILARITY.
 FT DISULFID 1358 1367 BY SIMILARITY.
 FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 185 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 794 794 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 799 799 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1009 1009 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1010 1010 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1019 1019 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1183 1183 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1266 1266 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1300 1300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 258 258 S -> SDEED (in isoform 2).
 FT VARSPLIC 480 487 /FTid=VSP_050035.
 FT Missing (in isoform 2 and isoform 3).
 FT /FTid=VSP_050036.
 FT Q -> K (IN REF. 2).
 FT CONFLICT 226 226 SLKT -> KPON (IN REF. 3).
 FT CONFLICT 607 610 L -> M (IN REF. 3).
 FT CONFLICT 634 634
 FT SEQUENCE 1529 AA; 169869 MW; 5D19CC5E7FD461BA CRC64;
 Query Match 10.4%; Score 397.5; DB 1; Length 1529;
 Best Local Similarity 29.7%; Pred.No. 1.6e-17;
 Matches 108; Conservative 58; Mismatches 141; Indels 57; Gaps 18;
 QY 158 HC---PTCRDRPCQGGQCHDSSESSVVCYCPAGFTGSRCEHSQL-----HCHPEAC 207
 Db 1072 HCDIDFDDCCQNKCKNGAHCCTDA-VNGYTCICPEYSGLFCFEPFPMVLPRTSPCDNFDC 1130
 QY 208 GPDATC---VNRPDGRGYTCRHLGRSLRCBEGVYVTPSLSGAGSYLALPAL-----T 259
 Db 1131 QNGAQCIVRINEP-----ICQCLPGYQEKCEKLVSV---NFINKESYLOTPSAKVRPQT 1182
 QY 260 NTHHELRLDVEFEKPLAPDGVLLVFGSKSPVEDFVSLAMVGGHLEFRYELGSGLA-VLRS 318
 Db 1183 NITLQIATDED-----SGILLYKGDK-----DHIAVELYGRVRSYDTGSHSPASAIYS 1231
 QY 319 AEPALGRHWRSVBRLNKDGSLRVNGRPPVLRSSPGKSOGLNHLTLLYLGGV-----P 373
 Db 1232 VETINDGNFHVLLALDQSLVSDGNGPKIITNLSKOSTLNFDSPLYVGGMPKSNVA 1291
 QY 374 SVPLSPATMGAHFRCCVGEVSNVGRKLDLTYFSLGSGIQCGQDSSPCERQPCOHGATC 433
 Db 1292 SLRQAPGQN-GTSPFGCIRNLYNSELQDFQKVPMTGILPGC---EPCHKVCAHG-TC 1346
 QY 434 MPAGEYEFQCLCRDGFKGDLCHEHEN-PCQLREPCCLHGHTCQ-----GTRCLCLPFGSGP 487
 Db 1347 QBSQAGTFCBCEQSWMFLCDQRINDPC-LGNKCVH-GTCLPINAFYSYCKCLEHGGV 1404
 QY 488 RCQQ 491
 Db 1405 LCDE 1408
 RESULT 15
 CLR3_HUMAN STANDARD; PRT; 3312 AA.
 AC Q9NYQ7; O75092;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cadherin EGF LAG seven-pass G-type receptor 3 precursor (Flamingo
 DE homolog 1) (hfmil) (Multiple epidermal growth factor-like domains 2)
 DE (Epidermal growth factor-like 1).
 GN CELSR3 OR CDHF11 OR FM11 OR EGF11 OR MEGF2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=20202599; PubMed=10716726;
 RX Wu Q., Maniatis T.;
 RT "Large exons encoding multiple ectodomains are a characteristic
 RL feature of protocadherin genes."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
 RN [2]
 RP SEQUENCE OF 1954-3312 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98360089; PubMed=9693030;

RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
 RT "Identification of high-molecular-weight proteins with multiple
 EGF-like motifs by motif-trap screening.";
 RL Genomics 51:27-34(1998).
 CC -!- FUNCTION: Receptor that may have an important role in cell/cell
 CC signaling during nervous system formation.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
 CC -!- SIMILARITY: Contains 9 cadherin domains.
 CC -!- SIMILARITY: Contains 8 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 laminin G-like domains.
 CC -!- SIMILARITY: Contains 1 laminin EGF-like domain.
 CC -!- SIMILARITY: Contains 1 GPS domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR ENBL; AE231023; RAFP61929.1; --
 DR ENBL; AS011536; EAA32464.1; --
 DR HSSP; P00740; LEDM.
 DR Genew; HGNC:3230; CELSR3.
 DR MIM; 604264; --
 DR GO; GO:0005198; F:structural molecule activity; NAS.
 DR InterPro; IPR000152; Asx hydroxyl_S.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR008985; ConA_like_1ec_g1.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR000832; GPCR secretin.
 DR InterPro; IPR001879; horwn_receptor.
 DR InterPro; IPR006210; LEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR000203; PKD_cyb_rich.
 DR Pfam; PF00002; 7tm_2; 1.
 DR Pfam; PF00028; cadherin; 9.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF01825; GPS; 1.
 DR Pfam; PF02793; HRM; 1.
 DR Pfam; PF00054; laminin_G; 2.
 DR PRINTS; PR00205; CADHERIN.
 DR PRINTS; PR00011; EGFLAMININ.
 DR PRINTS; PR00249; GPCRSECRETIN.
 DR SMART; SM00112; CA; 9.
 DR SMART; SM00181; EGF; 6.
 DR SMART; SM00303; GPS; 1.
 DR SMART; SM00008; Horwmr; 1.
 DR SMART; SM00282; Lamg; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00232; CADHERIN_1; 7.
 DR PROSITE; PS0268; CADHERIN_2; 8.
 DR PROSITE; PS00022; EGF_1_6_2; 4.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS00025; EGF_3; 6.
 DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
 DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
 DR PROSITE; PS00227; G_PROTEIN_RECEP_F2_3; 1.
 DR PROSITE; PS0261; G_PROTEIN_RECEP_F2_4; 1.
 DR PROSITE; PS02021; GPS; 1.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 2.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 DR G-protein coupled receptor; transmembrane; Glycoprotein;
 KW EGF-like domain; Calcium-binding; laminin EGF-like domain; Repeat;
 KW Developmental protein; Hydroxylation; Signal.
 FT SIGNAL 1 32
 FT CHAIN 33 3312
 FT CADHERIN EGF LAG SEVEN-PASS G-TYPE
 FT RECEPTOR 3.

FT	DOMAIN	2540	33	EXTRACELLULAR (POTENTIAL)	FT	CARBOHYD	2474	2474	N-LINKED (GLCNAC . .) (POTENTIAL)
FT	TRANSMEM	2541	2561	1 (POTENTIAL)	FT	CARBOHYD	2506	2506	N-LINKED (GLCNAC . .) (POTENTIAL)
FT	TRANSMEM	2562	2573	CYTOPASMIC (POTENTIAL)	FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2)
FT	TRANSMEM	2573	2593	2 (POTENTIAL)	FT	SEQUENCE	3312	3312	AA; 358251 MW; BEC208703651A4AS CRC64;
FT	TRANSMEM	2594	2601	EXTRACELLULAR (POTENTIAL)					
FT	TRANSMEM	2602	2622	3 (POTENTIAL)					
FT	TRANSMEM	2623	2643	CYTOPASMIC (POTENTIAL)					
FT	TRANSMEM	2644	2664	4 (POTENTIAL)					
FT	TRANSMEM	2665	2681	EXTRACELLULAR (POTENTIAL)					
FT	TRANSMEM	2682	2702	5 (POTENTIAL)					
FT	TRANSMEM	2703	2725	CYTOPASMIC (POTENTIAL)					
FT	TRANSMEM	2726	2746	6 (POTENTIAL)					
FT	TRANSMEM	2747	2753	EXTRACELLULAR (POTENTIAL)					
FT	TRANSMEM	2754	2774	7 (POTENTIAL)					
FT	TRANSMEM	2775	3312	CYTOPASMIC (POTENTIAL)					
FT	TRANSMEM	3312	3312	CADHERIN 1.					
FT	TRANSMEM	3326	433	CADHERIN 2.					
FT	TRANSMEM	434	545	CADHERIN 3.					
FT	TRANSMEM	546	651	CADHERIN 4.					
FT	TRANSMEM	652	756	CADHERIN 5.					
FT	TRANSMEM	757	858	CADHERIN 6.					
FT	TRANSMEM	859	961	CADHERIN 7.					
FT	TRANSMEM	962	1067	CADHERIN 8.					
FT	TRANSMEM	1068	1169	CADHERIN 9.					
FT	TRANSMEM	1170	1265	EGF-LIKE 1, CALCIUM-BINDING.					
FT	TRANSMEM	1375	1433	EGF-LIKE 2, CALCIUM-BINDING.					
FT	TRANSMEM	1435	1471	EGF-LIKE 3, CALCIUM-BINDING.					
FT	TRANSMEM	1475	1514	LAMININ G-LIKE 1.					
FT	TRANSMEM	1515	1719	EGF-LIKE 4, CALCIUM-BINDING.					
FT	TRANSMEM	1722	1758	LAMININ G-LIKE 2.					
FT	TRANSMEM	1764	1944	EGF-LIKE 5, CALCIUM-BINDING.					
FT	TRANSMEM	1946	1982	EGF-LIKE 6, CALCIUM-BINDING.					
FT	TRANSMEM	1983	2020	EGF-LIKE 7, CALCIUM-BINDING.					
FT	TRANSMEM	2021	2053	EGF-LIKE 8, CALCIUM-BINDING.					
FT	TRANSMEM	2055	2090	LAMININ EGF-LIKE.					
FT	TRANSMEM	2096	2131	GPS.					
FT	TRANSMEM	2477	2529	BY SIMILARITY.					
FT	TRANSMEM	1379	1390	BY SIMILARITY.					
FT	TRANSMEM	1384	1421	BY SIMILARITY.					
FT	TRANSMEM	1423	1432	BY SIMILARITY.					
FT	TRANSMEM	1439	1450	BY SIMILARITY.					
FT	TRANSMEM	1444	1459	BY SIMILARITY.					
FT	TRANSMEM	1461	1470	BY SIMILARITY.					
FT	TRANSMEM	1479	1490	BY SIMILARITY.					
FT	TRANSMEM	1484	1500	BY SIMILARITY.					
FT	TRANSMEM	1502	1513	BY SIMILARITY.					
FT	TRANSMEM	1726	1737	BY SIMILARITY.					
FT	TRANSMEM	1731	1746	BY SIMILARITY.					
FT	TRANSMEM	1748	1757	BY SIMILARITY.					
FT	TRANSMEM	1950	1961	BY SIMILARITY.					
FT	TRANSMEM	1955	1970	BY SIMILARITY.					
FT	TRANSMEM	1972	1981	BY SIMILARITY.					
FT	TRANSMEM	1985	1996	BY SIMILARITY.					
FT	TRANSMEM	1990	2008	BY SIMILARITY.					
FT	TRANSMEM	2010	2019	BY SIMILARITY.					
FT	TRANSMEM	2027	2040	BY SIMILARITY.					
FT	TRANSMEM	2042	2052	BY SIMILARITY.					
FT	TRANSMEM	2059	2074	BY SIMILARITY.					
FT	TRANSMEM	2061	2077	BY SIMILARITY.					
FT	TRANSMEM	2079	2089	HYDROXYLATION (POTENTIAL)					
FT	TRANSMEM	1963	1963	N-LINKED (GLCNAC . .) (POTENTIAL)					
FT	TRANSMEM	632	632	N-LINKED (GLCNAC . .) (POTENTIAL)					
FT	TRANSMEM	847	847	N-LINKED (GLCNAC . .) (POTENTIAL)					
FT	TRANSMEM	1182	1182	N-LINKED (GLCNAC . .) (POTENTIAL)					
FT	TRANSMEM	1222	1222	N-LINKED (GLCNAC . .) (POTENTIAL)					
FT	TRANSMEM	1317	1317	N-LINKED (GLCNAC . .) (POTENTIAL)					
FT	TRANSMEM	1327	1327	N-LINKED (GLCNAC . .) (POTENTIAL)					
FT	TRANSMEM	1649	1649	N-LINKED (GLCNAC . .) (POTENTIAL)					
FT	TRANSMEM	1713	1713	N-LINKED (GLCNAC . .) (POTENTIAL)					
FT	TRANSMEM	1770	1770	N-LINKED (GLCNAC . .) (POTENTIAL)					
FT	TRANSMEM	2053	2053	N-LINKED (GLCNAC . .) (POTENTIAL)					
FT	TRANSMEM	2177	2177	N-LINKED (GLCNAC . .) (POTENTIAL)					
FT	TRANSMEM	2196	2196	N-LINKED (GLCNAC . .) (POTENTIAL)					
FT	TRANSMEM	2386	2386	N-LINKED (GLCNAC . .) (POTENTIAL)					

Search completed: March 9, 2004, 17:19:43
 Job time : 22.0517 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:15:08 ; Search time 61.7389 Seconds
(without alignments)
3602.917 Million cell updates/sec

Title: US-10-006-011a-3
Perfect score: 3825
Sequence: 1 EIKITRPSADGMLLYNGQ.....QLDLQHRAGAGANTRPCPS 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mmc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	901.5	23.6	2026	4	O00468
2	840	22.0	775	4	Q8N7Y0
3	820	21.4	1009	11	QBGGP3
4	812.5	21.2	708	5	Q9NFS9
5	812.5	21.2	1035	5	Q9NEG1
6	812.5	21.2	4117	5	Q8IRV9
7	812.5	21.2	4179	5	Q9W4Y4
8	812.5	21.2	4223	5	Q8WPN3
9	812.5	21.2	4228	5	Q8IRV8
10	703	18.4	488	4	Q961C1
11	664.5	17.4	1361	5	Q9V714
12	664.5	17.4	1361	5	Q9NGV2
13	628	16.4	463	4	Q8NAL2
14	581.5	15.2	822	11	Q8RU56
15	576.5	15.1	1039	5	Q9V0B1
16	511	13.4	925	5	Q9U7E8

ID	Accession	Score	Match	Length	DB ID	Description
17	479	12.5	350	11	Q8K3Z6	Q8K3Z6 mus musculus
18	464.5	12.1	478	5	Q66D6	Q66D6 ciona intes
19	420.5	11.0	1458	11	Q8CJG8	Q8CJG8 rattus norv
20	419.5	11.0	1531	11	Q88279	Q88279 rattus norv
21	412.5	10.8	1531	11	Q9WVB5	Q9WVB5 mus musculus
22	412.5	10.8	1557	11	Q80TR4	Q80TR4 mus musculus
23	411.5	10.8	1530	11	Q9WUG5	Q9WUG5 rattus norv
24	407.5	10.7	1474	11	Q8CJG9	Q8CJG9 rattus norv
25	407	10.6	1534	4	Q75093	Q75093 homo sapien
26	407	10.6	1518	4	Q9UIL7	Q9UIL7 homo sapien
27	402.5	10.5	1176	11	Q80Y87	Q80Y87 mus musculus
28	402.5	10.5	1553	11	Q8CH86	Q8CH86 mus musculus
29	394.5	10.3	1523	4	Q75094	Q75094 homo sapien
30	393.5	10.3	517	5	Q8NRJ7	Q8NRJ7 drosophila
31	390	10.2	1512	13	Q9DE36	Q9DE36 brachydanio
32	387.5	10.1	1025	11	Q9Z166	Q9Z166 mus musculus
33	387.5	10.1	1521	11	Q9R1B9	Q9R1B9 mus musculus
34	387	10.1	1523	11	Q88280	Q88280 rattus norv
35	385.5	10.1	950	13	Q90Z44	Q90Z44 gallus gall
36	385.5	10.1	1392	4	Q8IUE3	Q8IUE3 homo sapien
37	385	10.1	1523	11	Q9WVB4	Q9WVB4 mus musculus
38	380.5	9.9	313	13	Q9YHFO	Q9YHFO rana pipien
39	377.5	9.9	1529	13	Q7ZX12	Q7ZX12 xenopus lae
40	375	9.8	738	13	Q90Z45	Q90Z45 gallus gall
41	370.5	9.7	1100	11	Q8CCT8	Q8CCT8 mus musculus
42	367	9.6	1515	13	Q9DE37	Q9DE37 brachydanio
43	363	9.5	1837	5	Q9VCZ9	Q9VCZ9 drosophila
44	362.5	9.5	4307	5	Q19319	Q19319 caenorhabdi
45	361.5	9.5	792	13	Q90Z43	Q90Z43 gallus gall

ALIGNMENTS

RESULT 1
O00468 PRELIMINARY; PRT; 2026 AA.
AC O00468;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE AGRIN precursor (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Rutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Groffen A.J.A., Buskens C.A.F., Van Kuppevelt T.H.M.S.M.,
RA Veerkamp J.H., Monnens L.A.H., van den Heuvel L.P.W.J.;
RL Eur. J. Biochem. 0:0-0(1998).
RN [2]
RP SEQUENCE OF 1-153 FROM N.A.
RX MEDLINE=9624170; PubMed=8617505;
RA Lennon G., Auffray C., Polymeropoulos M., Soares M.B.;
RA "The I.M.A.G.E. Consortium: an integrated molecular analysis of
RT genomes and their expression.";
RL Genomics 33:151-152(1996).
RN [3]
RP SEQUENCE OF 1-153 FROM N.A.
RA Denzer A.J., Brandenberger R., Gesemann M., Chiquet M., Ruegg M.A.;
RL J. Cell Biol. 0:0-0(0).
DR EMBL; AF016903; AAC39776.1; -
DR EMBL; U84406; AAB52917.1; -
DR HSSP; P00740; 1EDM.
DR Genew; HGNC:329; AGRN.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR004850; Agrin_NcA.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR003645; FcLN.
DR InterPro; IPR002350; Kazal.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.

DR InterPro; IPR000082; SEA domain.
 DR InterPro; IPR008993; TIMP like.
 DR InterPro; IPR001455; UFP0033.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00050; Kazal; 9.
 DR Pfam; PF00053; laminin_EGF; 2.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF03146; NCA; 1.
 DR Pfam; PF03390; SEA; 1.
 DR PRINTS; PRO0011; EGF_LAMININ.
 DR SMART; SM00180; EGF_Lam; 2.
 DR SMART; SM00274; FOLN; 5.
 DR SMART; SM00280; KAZAL; 9.
 DR SMART; SM00282; LamG; 3.
 DR SMART; SM00200; SEA; 1.
 DR PROSITE; PS00022; EGF_1; 6.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
 DR PROSITE; PS00024; SEA; 1.
 DR PROSITE; PS01148; UFP0033; 1.
 DR EGF-like domain; Laminin EGF-like domain; Signal.
 KW NON_TER 1
 FT SIGNAL <1 10 POTENTIAL.
 FT CHAIN 11 2026 AGRIN.
 FT SEQUENCE 2026 AA; 212881 MW; 4AB0EE710CD4B8EF CRC64;
 SQ
 Query Match 23.6%; Score 901.5; DB 4; Length 2026;
 Best Local Similarity 32.2%; Pred. No. 3.2e-58;
 Matches 242; Conservative 93; Mismatches 266; Indels 151; Gaps 22;
 QY 2 IKITFRPDSADGMLLYNGQKRVGSPFTNLANKQPPFISGLVGRPEPRFDAGSGMATIR 61
 DB 1377 LALEFRALPEQGLLYNGNA-----RGKDFLALALDDGRVQLRFRFTGSGPAVLT 1425
 QY 62 HPTPLALGHFHTVTLRLSTOGSLIVGDLAPVNGTSQKFGQLDNLNELYLGGYDP--Y 118
 DB 1426 SAVPVEPGQWHLELSRHWRRGTLSDVGETFVLGSPGTDGLNDLDFVGGVPEDQAA 1485
 QY 119 GAIPKAGLSGGTGVCRELRIQGEETIVFDLNL-----TAHGISHC--PTCRDRPCQNG 170
 DB 1486 VALERFVAGLRGRCIRLDVNNQRL--ELGIGPAAATRGSGVGCEDHPCLNPNCHGG 1542
 QY 171 GQCHDSESSYVVCVCPAGTSGRCEHSQALHCHPEACGPDATCVNRPPGRGYTCRCHLGR 230
 DB 1543 APCQNLSEAGRFHQCCPPRGVPTCADEKS--PCQNPCHGAAPCRVLPPEG-GAQCCEPLGR 1600
 QY 231 SGLRCEGVTVTPSLSGAGSYLALPALNTNTHHEL-----LDVEFKPLAP 276
 DB 1601 EGTFCQ-----TASGQDGGSPFLA-DFNGFSHLELRGLHTFPARDLGERKMALEVVELARGP 1654
 QY 277 DGVLLFSGKSGVPEDFVSLAMVGGHLEFRYELGSLAVDRSAEPLALGRHVRVSAERLN 336
 DB 1655 SGLLYNGQKTDGKGFVSLARDRELEFRYDLGKGAAVIRSERPVTLGAWTRVSLERNG 1714
 QY 337 KDGSLRVNGRPRVLRSPGKSOGLNLHTLLYLGGVPSVPLSPATMNSAHFRGCVGEYSV 396
 DB 1715 RKALAVGDGPRVVLGSESPVPHVTLNKEPLYVGGAPDFSKLARAANAASVSGFDGAIOLVSL 1774
 QY 397 NGRKL-----DLTVSFLGSGIQCYDSSPCER---QPCQHGATCMPAGEVEFCQ 443
 DB 1775 GGRQLLTPHVLQRVDVT-SFAG-----HPCTRASGHFCLNGASCVP----- 1815
 QY 444 LCRDGFKDLCEHENPCQLREPLCGHTGTCQTRCLCLPFGSPRCQOQSGHGIAESDWH 503
 DB 1816 -----REAY-----VCLCPGFGSGPCKEGL----- 1837
 QY 504 LEGSGNDAPQGYAVPHDGFALFGVHFSRSLPEVPET-----LELVRSIATSG 555
 DB 1838 VEKSAG-----DVIDTLAFDGRTEVEYLVNATVESKALQSNHPELSLRTEATQG 1885
 QY 556 LLLWQGVVEVAGQKDFISLQDGHVFRYQLGSGEARLVSEDEPINDGEHWRHTVALRE 615

DB 1886 LVLMS-----GKATERADYVALAIVDGHQLQLSNLGSQPVLRSTVPVNTNRMLRVVAHRE 1941
 QY 616 GRGSLQVDCBELYSGRSPGNVAVNAKGVYIGGAPD--VATLTGGRFSSGITGVKNL 673
 DB 1942 QREGSLQVGNAPVTGSSPLGATQLDTGALWIGGLPELPGVGFALPKAYGTGFVGCLEADV 2001
 QY 674 VLHSARPAGAPPQPLDQHQRAQAGANTRFPCPS 705
 DB 2002 VVGR-----HPLHLLEDAVTKPELRPCPT 2025
 PRELIMINARY; PRT; 775 AA.
 Q8N7Y0
 ID Q8N7Y0
 AC Q8N7Y0;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein FLJ40230.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
 RA Orsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
 RA Marukawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
 RA Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
 RA "NEDO human cDNA sequencing project";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK097549; BAC05096.1;
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR InterPro; IPR008985; ConA_like_lec_gl.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; LEGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00054; laminin_G; 3.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SM00282; LamG; 3.
 DR PROSITE; PS00022; EGF_1; 3.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
 KW Hypothetical protein; EGF-like domain.
 SQ SEQUENCE 775 AA; 84800 MW; 61A81294F0204ED2 CRC64;
 Query Match 22.0%; Score 840; DB 4; Length 775;
 Best Local Similarity 29.6%; Pred. No. 3.6e-54;
 Matches 214; Conservative 107; Mismatches 257; Indels 144; Gaps 20;
 QY 1 EKITFRPDSADGMLLYNGQKRVGSPFTNLANKQPPFISGLVGRPEPRFDAGSGMATI 60
 DB 176 QITLFRFAEAEDGILLYCGENE-----HGRGDFMSLAIRRSIQFRFNGCTGVAVI 226
 QY 61 RHPTPLALGHFHTVTLRLSTOGSLIVGDLAPVNGTSQKFGQLDNLNELYLGGYDP-YG 119
 DB 227 VSETKILGWHITVLRDGLNGLLQLNNGTPTVQSGQGYSKITFRTPLYLGGAPSAW 286
 QY 120 AIPKAGLSGGTGVCRELRIQGEETIVFDLNL-----LTAHGISHCPT--CRDRPCQNG 170
 DB 287 LVRATGNTNRGQGVQSLAVNGRRI---DMRPWPLGKALSGADYVGECSGICDEASCIHG 343
 QY 171 GQCHDSESSYVVCVCPAGTSGRCEHSQALHCHPEACGPDATCVNRPPGRGYTCRCHLGR 230
 DB 344 GTCTAIKADSVIICLCPLOF----- 362
 QY 231 SGLRCEGVTVTPSL-SGAGSYLALPALNTNTHHEL-----LDVEFKPLAPDGVLLFS- 283

Db 363 KGRHCEAFTLTIQFRESLSYAAATPWPLEPQHVLSFMEFIITFRDSDGCVLLSYDT 422
 Qy 284 GKSGPVEDFVSLAMVGGHLEFRYELGSLAVLSAEPALALGRWHRVSAERLNKDGSLRV 343
 Db 423 GSK-----DFLSINLAGGHVFRFCGSGTGLRASEDELTLGNWHELRSRTAKNGILQV 477
 Qy 344 NGRPVLSRSPKSGQLNLHTLLYLGVEPSVPLSPATNMSAHRFGCVGVNKGKLDL 403
 Db 478 DKQKIVGMAEGGFTQIKNTDIFIGVNPVDDVKNSGVLPKSPGSIQKILINDRIHV 537
 Qy 404 TVSFILGSGIGCYDSSPCERQPCOHGATCMPAGEYEFQCLCRDPGFKGDLCEHEENPCQL 463
 Db 538 KHPF--TSGVNVVNAAHPCVRAPCAHGSCRPKE-GYDCDPLGFEF----- 582
 Qy 464 REPCLHGTCQGT--RCLPLGSPRCQSGHGIAESDWHLEGGSGNDAPGOYGFH 521
 Db 583 -----LH-----CQKAIIEAIEIPIQFIG-----RSYLYDYN 608
 Qy 522 DDGFLAFPHVFRSRLPEVETIELEVRTSTASGILLWQGVGEGAGQKDFISLGLQDG 581
 Db 609 PDLKRVSG---SRS-----NVFMRFKTKAKDGLLNRG--DSPMRPNSDFISLGLRDLG 657
 Qy 582 HLVFRYQLGSEARLVSDPINDGEWHRVTALREGRSIQVDSBELYSGRSPGNVAVN 641
 Db 658 ALVFSYNLGSGVASIMVANGSFDNDRWHRVYKVRDQSGKIVDDYGARTGKSPGMRLQN 717
 Qy 642 AKGSVYVIGAPDVAITLGTGRFSSGITGVKNLVLSARFPAPPQPLDLQHRAGANTR 701
 Db 718 INGALYVGMKEIALHTNRQYMRGLVGCLSHFTLST-----DYHISLVEDAVDQKIN 770
 Qy 702 PC 703
 Db 771 TC 772

RESULT 3
 Q8BGP3 PRELIMINARY; PRT; 1009 AA.
 ID C8BGP3;
 AC Q8BGP3;
 DT 01-NAR-2003 (TrEMBLrel. 23, Created)
 DT 01-NAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical EGF-like domain.
 GN AU040377.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=c57BL/6J; TISSUE=Forelimb, Skin, Testis, and Thymus;
 RA MEDLINE=22354683; Pubmed=12466851;
 RX THE FANTOM Consortium,
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 DR EMBL; AK031136; BAC27271.1; -
 DR EMBL; AK033332; BAC28235.1; -
 DR EMBL; AK037223; BAC29762.1; -
 DR EMBL; AK041546; BAC30982.1; -
 DR MGD; MGI:2146149; AU040377.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR InterPro; IPR008985; ConA_like_lec_1.
 DR InterPro; IPR00742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR06209; EGF_like.
 DR InterPro; IPR003962; FnIII subd.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR008957; FN_III-like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin_G.

Pfam; PF00008; EGF; 2.
 Pfam; PF00041; fn3; 2.
 Pfam; PF00064; Laminin_G; 3.
 DR PRINTS; PR00014; ENTPEFII.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SM00060; FN3; 2.
 DR SMART; SM00282; LamG; 3.
 DR PROSITE; PS00022; EGF_1; 3.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS50025; Lam_G_DOMAIN; 3.
 KW Hypochemical protein.
 SQ SEQUENCE 1009 AA; 109834 MW; 80DD954F255C102B CRC64;
 Query Match 21.4%; Score 820; DB 11; Length 1009;
 Best Local Similarity 29.3%; Pred. No. 1.6e-52;
 Matches 211; Conservative 109; Mismatches 261; Indels 138; Gaps 19;
 Qy 1 EIKITRPSADGMLYNGKRVSPGPTLANRQDPFISGLVGGPFRFPAAGSMATII 60
 Db 410 QVTLFRAEADGLLLYCGESE-----HGRDFMSLALRRSLHRFRNCTGIAII 460
 Qy 61 RHPTPLALGHFHTVTLRLSITQGLSIVDGLAPVNGTSQKFGQGLDNLNELYLGGYDP-YG 119
 Db 461 ISETKIKLGAHWVTVLYRDGLNGMLQNNGTPTVQSGOQYKITERPTPLYLGGAPSAW 520
 Qy 120 AIPKAGLSSGFIGCVRELRIOGEEIVFHDLN-----LTAHGISHCPT--CRDRPCQNG 170
 Db 521 LVRAATGNRFQCVQSLVNGKII---DMRPWPLKALNGADVGECCSGICDEASCIHG 577
 Qy 171 GQCHDSSESYVVCVAGFTGSRCSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
 Db 578 GTCAAIKADSYICLPLGFRHCE-----DAPALTIPO-----FR 613
 Qy 231 SGLRCEGVTVTPFSLSGAGSYLALPALTWHHEL---RLDVEFKLAPDGVLLFS---G 284
 Db 614 ESLR-----SYAATPWPLEPQHVLSFMEFIITFRDSDGCVLLSYDTG 657
 Qy 285 GKSGPVEDFVSLAMVGGHLEFRYELGSLAVLSAEPALALGRWHRVSAERLNKDGSLRVN 344
 Db 658 SK-----DFLSINLAGGHVFRFCGSGTGLRASEDELTLGNWHELRSRTAKNGILQVD 712
 Qy 345 GGRPVLSRSPKSGQLNLHTLLYLGVEPSVPLSPATNMSAHRFGCVGVNKGKLDLT 404
 Db 713 KQVVEGMARGGFTQIKNTDIFIGVNPVDDVKNSGILHPFSGSIQKILINDRIHV 772
 Qy 405 YSFLGSGIGCYDSSPCERQPCOHGATCMPAGEYEFQCLCRDPGFKGDLCEHEENPCQLR 464
 Db 773 HDF--TSGVNVVNAAHPCVRAPCAHGSCRPKE-GYECDCPLGFEGLNCKAI----- 823
 Qy 465 EPLHLGGTCQGTCLPLGSPRCQSGHGIAESDWHLEGGSGNDAPGOYGFHDDG 524
 Db 824 -----IEAIEIPIQFIRSYLYTNDPNILK---RVSGSRSN-----A 856
 Qy 525 FLAPPHVFRSRLPEVETIELEVRTSTASGILLWQGVGEGAGQKDFISLGLQDGLHV 584
 Db 857 FMRF-----KYTKADGLLNRG--DSPMRPNSDFISLGLRDLGALI 894
 Qy 585 FRYQLGSEARLVSDPINDGEWHRVTALREGRSIQVDSBELYSGRSPGNVAVNAKG 644
 Db 895 PSYNLGSGVASIMVANGSFDNDRWHRVYKVRDQSGKIVDDYGARTGKSPGLMRLQNG 954
 Qy 645 SVYTGAPDVAITLGTGRFSSGITGVKNLVLSARFPAPPQPLDLQHRAGANTRPC 703
 Db 955 ALVYVGMKEIALHTNRQYMRGLVGCLSHFTLST-----DYHISLVEDAVDQKINTC 1006
 RESULT 4
 Q8NFS9 PRELIMINARY; PRT; 708 AA.
 ID Q8NFS9
 AC Q8NFS9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

```

DB 478 CENGGSCSEQEDVAVVCSQPFGRGHCQE-----HLQLG-----FNASPRG 518
QY 523 DGLFAPPGHVFRSRLPEVPEFTELEVRVSTAGSLLWGVGEAGQKDFISLGLQDGH 582
DB 519 DGYVELNRSHFPALAEQSYTSMGIVFTINKFNGLLFWMGQEAQEYTGQDFLAAAADVGY 578
QY 583 LVFRYQLGSGEARLVSDE-PINDGEMHRVTRALRGRGSIQVDGSELVSGRS-PGFNVAV 640
DB 579 VEYSMELDGEAEVIRNSDI.RVDNGERHVIARNDENTAILLEVD-RMLHSGETRPTSJKSM 637
QY 641 NAKGSYVICGADVDVATLTCGRSSGTCVKMLLHLSRPPGAPPQPLDLQHRAGANT 700
DB 638 KLFQNVFVCGGAPDLEVTFTGFRYKHLNGCI--VVVEGETVG-----QINLSSAAVNGVNA 690
QY 701 RPPCS 705
DB 691 NVCPA 695

RESULT 5
Q9NEG1 PRELIMINARY; PRT; 1035 AA.
AC Q9NEG1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE EG:BACR25B3.11 protein.
OS TROL OR EG:BACR25B3.11 OR CG7981.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP MURPHY L., HARRIS D., BARRELL B.;
RA "Sequencing the distal X chromosome of Drosophila melanogaster.";
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; ALI38972; CAB72284.1; -
DR HSSP; P00740; 1EDM.
DR FlyBase; FBgn001402; trol.
DR GO; GO:0003676; F.nucleic acid binding; IEA.
DR InterPro; IPR006985; ConA_like Lec_gl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR007110; Ig-Like.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00008; EGF_3.
DR Pfam; PF00047; Ig; 2.
DR Pfam; PF00054; Laminin_G; 3.
DR Pfam; PF00057; ldl_recept_a; 1.
DR SMART; SM00181; EGF_2.
DR SMART; SM00408; IGC2; 2.
DR SMART; SM00282; LamG; 2.
DR SMART; SM00192; LDLa; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS50835; IG_LIKE; 2.
DR PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW EGF-like domain; Immunoglobulin domain.
SQ SEQUENCE 1035 AA; 114044 MW; 413DE7EDD3ABE711 CRC64;

```

```

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Perlecan domain V (Fragment)
GN TROL OR EG:BACR25B3.11 OR CG7981.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Friedrich M.V.K., Schneider M., Timpl R., Baumgartner S.;
RT "Perlecan domain V of Drosophila melanogaster: Sequence, recombinant
analysis and tissue expression.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AU271462; CAB70094.1; -.
DR HSSP; P00740; 1EDM.
DR FlyBase; FBgn001402; trol.
DR GO; GO:0003676; F.nucleic acid binding; IEA.
DR InterPro; IPR006985; ConA_like Lec_gl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00008; EGF; 3.
DR Pfam; PF00054; Laminin_G; 3.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00282; LamG; 3.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW EGF-like domain.
FT NON TER 1
SQ SEQUENCE 708 AA; 77786 MW; 7B67CE290642B98 CRC64;

Query Match 21.2%; Score 812.5; DB 5; Length 708;
Best Local Similarity 30.2%; Pred. No. 3.6e-52;
Matches 219; Conservative 121; Mismatches 304; Indels 81; Gaps 22;

QY 3 KITFRPDSADGMLLYNQKRVFQSPFNLANRPFDFISFLVGVGPEFRFPDAGSGMATIRH 62
DB 30 ELTFRENGDGLLFFNGQTRGSG-----DYIALSLKDRYAEFRDFGCKPMLVRA 79
QY 63 PTPALGHHTVTLRLSLQGSLLIUGDLAPVNGTSQKQGGDLNLEELVLSGYPDYGAIP 122
DB 80 EEPALNNEHWTVRSRFRKRDGYIQVDSQHPVAPPTLQQIPQLDIEDLYIGVGNWELLIP 139
QY 123 KAGLSS--GFIGVRELRIOGEEIVFHDNLMLTAHSHGCHPTCRDRPCQNGGCHDSSES-S 179
DB 140 ADAVSQQVGFVGCISRLTLQGRIVELIIRAKYKEGTTDCRCAQGFQCKNGVCLSEOTBQ 199
QY 180 SYVCVCPAGTSGRCSHSAHQHCHPEACGPDATCVNRPDGRGVTCECHLGRSLACEEYV 239
DB 200 AYTICICPGWTRDCA-IEGTOCTPGYCGA-GRCENTEN--DMECLCLPNSRSGDRQYNE 255
QY 240 TVTTPSLSGAGSYLA--LPALNTWHEHLRDLVEFKPLA-PDGVLLPFGSKGKGVVDFVYS 295
DB 256 ILNEHSLNFNGNSFAAYGTPKVT-----KVNITLSVRPASLEDSVILYTAESTLPSGDYLA 311
QY 296 LAMVGGHLEFRYELGSL--AVLRSAAPLALGRHWRSARLNKQGLSRVNGRPFVLRSS 353
DB 312 LVLRGCHAEILLNTAARLDVWVRSAPLPLNWRTRIEIRRRLEGILRVGDGPERKAKA 371
QY 354 RKSQGLNHLTLVLSGVEPS-VPLSPATNMSAHRGCVGEVSVNGVKRLDLYTSLGSGQ 412
DB 372 PGSRIILSKTHLVGGVDSVTKVNRDVMITKFGDCISRLYNFQKPYNLLADIKDAAN 431
QY 413 IGQCYDSSPCBRQPCQHGATCMPAGEYE-----FOCLCRDGFKGDICEHEENPCQIREP 466
DB 432 IQSC-----GETNMIGDSDSNEPVPPEPTDVENELQPYWAPC-ASDP 477
QY 467 CLHSGTCQGITR- ---CLCLPFGSFRCCQSGHGIAESDWHLESGSGGNDAPQGYGAYFHD 522

```

Query Match 21.2%; Score 812.5; DB 5; Length 1035;
 Best Local Similarity 30.2%; Pred. No. 6e-52;
 Matches 219; Conservative 121; Mismatches 304; Indels 81; Gaps 22;

OY 3 KITRPSADGMLLVNGQRVGSPTLNANKRPPFISGLVGGPFRFRFAGSGMATIRH 62
 DB ELTRFPENGGLLLFNGQTRSG-----DYIALSLKDRYAEFRFPFGKPMLVRA 419

OY 63 PTPALGHFTHTVLLRSLTQGLSLVGLDLPVNGTSGKFKQGLDNLNEELLYLGGVDPYGAIP 122
 DB EEPALNEMHWVRSFRKSDYIQVDEGHVFAFTLQIQIPOLDLIEDLYIGVFNWELLP 479

OY 123 KAGLSS--GFTGCVRELRIGEEIVFHDLNLTANGISHCPTCRDRPCONGGQCHDSBS-S 179
 DB ADAVSQQVGFVGCISRLTQGRITVELIREAKYKRGITDCRPAQQCPQNGKVCLESQTEQ 539

OY 180 SYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCNVRPDGRGYTCRHLGSRGRCREGV 239
 DB 540 AYTCLIQEWTGRDCA-IEGTCQTFVCGA-GRENTEN--DMELCLFLNRSGRDRCCVNE 595

OY 240 TVTTPSLSGAGSYLA---LPALTNTHTHELRLDVEFKPLA-PDGVLLFSGGKSGVDFVFS 295
 DB ILNHSILNFKNGSFAAYGTPKVT----KVNITLSVRPASLEDSVILYTABSTLPSGDYLA 651

OY 296 LAMVGGHLEFRYELGSGI--AVLRSAPLALGRWHYSAERLNKDGSLRVNNGRPLRSS 353
 DB LVLRGGHAELLINTAARLDPPVVRSAEPLNWRTRIEIRRLGEGILRVGDGPERKAKA 711

OY 354 PKSQGLMLHTLLYLGGVEPS-VPLSPATNMSAHRFCVGVSVNGKRLDLYTYSFLSGOG 412
 DB PGSDRILSLKTHLYVGGYDRSVKVRDWNITKFGDGCISLRYNFKQFPVNLADIKDAAN 771

OY 413 IQCCYDSSPFCRQPCGHGATCMTAGEYE-----FQCLCRDGFKDLCEHENSCQLREP 466
 DB 772 IQSC-----GETNMIGDESDNEPPVPPPTPDVHNELOPYAMAFC-ASDP 817

OY 467 CLHGTGCOQTR----CLCLPFGSPRCQQSGHGHTAESDWHLESGNGDAPGQYGFPHD 522
 DB CENGGSCSEQEDVAVCVSPFGSGKHCQE-----HLQLG-----FNASFRG 858

OY 523 DGFLAPPGHVFRSRPEVPEETIELEVRSTASGLLWQVVEGRAGQKDFISLGLQDGH 582
 DB DGYVELNRSHFPALEQSYTSMGIYFTTKENGLLFWWQEGAGEEYTGDFIAAAVDGY 918

OY 583 LVFRYOLGSGEARLVSED--PINDGEWHRVTLALRGRRGSIQVDGEBELVSGRS-FGPNVAV 640
 DB VBYSMRLDGEAEVIRNSDIRVDNGERHIVIAKRDENTALEVD-RMLHSGETRPTSKSM 977

OY 641 NAKGSYVIGGAPDVATLTGFRSSGITGCKVNLVLSARPGAPPDPLDQHRAGAGANT 700
 DB KLPGNVFGGAPDLEVTGFRYKHNLANGCI--VVVEGETVG-----QINLSSAAVNGVNA 1030

OY 701 RCPGS 705
 DB 1031 NVCPA 1035

RESULT 6
 O8IRV9 PRELIMINARY; PRT; 4117 AA.
 AC O8IRV9;
 DT 01-MAR-2003 (TRMBLrel. 23, Created)
 DT 01-MAR-2003 (TRMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TRMBLrel. 25, Last annotation update)
 DE CG7981-FC
 GN TROL OR EG:BACR25B3.11 OR CG7981.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton R.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Aghayani A., An H.J., Andrews-Frankoch C., Saldwin D.,
 RA Ballaw R.M., Basu A., Bakendale J., Bayraktaroglu L., Besley E.M.,
 RA Beeson K.Y., Betchan M.R., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Dalush F., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hosten D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy J., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spierak R., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodagter, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A.; Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [2]

RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Bazzon J., An H., Baldwin D., Bazton J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hosten D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Fargacs V., Park S., Patel S., Pfeiffer B., Scheeler F.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradscky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smurniak P., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]

RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]

RA McIntosh T.C., Moy M., Murphy B., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "sequencing of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kamniker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Beriman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "annotation of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AJ487018; CAD34650.1; -
 DR EMBL; AE003424; AAN09079.1; -
 DR FlyBase; FBgn0001402; trol.
 DR GO; GO:0003577; F:DNA binding; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0006306; P:DNA methylation; IEA.
 DR InterPro; IPR001535; CS_DNA_meth.
 DR InterPro; IPR006985; Confl_likelike_rec_gl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR006209; EGF_likelike.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR007110; Ig_likelike.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00047; ig; 13.
 DR Pfam; PF00052; laminin_B; 3.
 DR Pfam; PF00053; laminin_EGF; 2.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF00057; ldl_recept_a; 23.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR PRODOM; PD003031; Laminin_B; 3.
 DR SMART; SM00180; EGF_Lam; 8.
 DR SMART; SM00409; IG; 13.
 DR SMART; SM00408; IGC2; 13.
 DR SMART; SM00281; LamB; 3.
 DR SMART; SM00282; LamC; 3.
 DR SMART; SM00192; LDLA; 23.
 DR PROSITE; PS00094; CS_MTASE_1; 1.
 DR PROSITE; PS00022; EGF_1; 10.
 DR PROSITE; PS01186; EGF_2; 6.
 DR PROSITE; PS50835; IG_LIKE; 12.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 3.
 DR PROSITE; PS01209; LDLRA_1; 20.
 DR PROSITE; PS50068; LDLRA_2; 23.
 DR PROSITE; PS00030; REM_RNP_1; 1.
 DR Immunoglobulin domain.
 KW SEQUENCE 4223 AA; 456919 MW; 31D7C2B5C9B1D2E0 CRC64;

RA Query Match 21.2%; Score 812.5; DB 5; Length 4233;
 RA Best Local Similarity 30.2%; Pred. No. 4e-51;
 RA Matches 219; Conservative 121; Mismatches 304; Indels 81; Gaps 22;

RA Preliminary; PRT; 4223 AA.
 ID Q8MPN3
 AC Q8MPN3
 DT 01-OCT-2002 (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DE Perlecan [CG7981-1.PD]
 GN TROL OR EG:BACK25B3.11 OR CG7981.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidae; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Voigt A.;
 RT "perlecan participates in proliferation activation of quiescent
 RT Drosophila neuroblasts";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; Pubmed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Borchen M.R., Bouck J., Brockstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA De Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Liu Z.,
 RA Liu X., Mattei B., McIntosh T.C., McCleod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodagter, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Smith H.O.,
 RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu S.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K.J., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

Qy	3	KITRFDSDMLLYAGQKRVPGSPPTNLANRPFDFISFGLVGRREFFRFDAGSGMATIRH	62
Db	3545	ELFRPENGDLLEFGNGTRGSG-----DYIALSKDRYAEFRDFRGKPMLVRA	3594
Qy	63	PTPLALGHFTWLLASLTCGSLIVGDLPVNGTSGCRKFOGDIIMBELYLOGYPYGAIP	122
Db	3595	EEFALNMEHTVRSRFRKDGXYQVDEPQVAFPTLQQLPPQLDLLEDLYIGGVNWEFLP	3654
Qy	123	KAGLSS--GFICGVRELRIQGEIIVPHDLNLTAKHGSHCTPDRPCPONGCQCHDSBS-S	179
Db	3655	ADAVSOQGVFGVCSIRLTLQGRVTELRVREAKYKGEITDRCPCAOQFCQKNGVCLSESQEQ	3714
Qy	180	SYVVCVPCAGTSGRCEHSQALCHHPAAGCPDTCVNRPDGRYTCRCHLGRSLRCEBGV	239
Db	3715	AYTICOPGWTRDCA-IEGTCTPVCQA-GRCENTEN--DMECCLPNSRGGKCYQNE	3770
Qy	240	TVTPSLSGAGSYLA---LPALNTTHTHELRDLVFKPLA-PDGVLVLFSGKSGPVEDFVS	295
Db	3771	ILNHSILNFKNQSAFAYGTPKVT---KVNITLSVRPASLEDSEVILYTAESTLPSGDYLA	3826
Qy	296	LAMVGHLEFRYELGSL--AVLRSAPLALGRWHRVSAERLNKLSLVNNGRVERLSS	353
Db	3827	LVRGHAELLINTRAARLDPVVRSAEPLPLNRWTRIBIRRLRIGGILRVGDDGPERKAKA	3886
Qy	354	PGKSQLNLTLLYLGGVEPS-VPLSPATMSAHRFGCVGVSYNGKRLDLDLTYSLGSGQ	412
Db	3887	PGSDRILLSLTKHLVYGGYDRSTVKNRDNVITKGFDDCISRLYNFQKFWNLLADIKDAAN	3946
Qy	413	IGOCYDSSPCRQPCQHGATCPAGEVE-----FQCCLRDFKGDLCHEHENPCQLREP	466
Db	3947	IQSC-----GETNMIGGDESDNPVPPPTPDPVHNELOPYAMAPC-ASDP	3992
Qy	467	CLRGGTCCQFTR-----CLCLFGFSRCCQGGHGAESDWHLGSGGNDAPGQYGAVFDH	522
Db	3993	CENGGSCSEQEDVAVCVCPFGSKHCOE-----HLQLG-----FNASFRG	4033
Qy	523	DGFLAPGHVFSRLEVEVTEILEVTRSTAGLILWQVVEAGEAGQKDFISLGLDQGH	582
Db	4034	DGVELNRSHFQALQSVTSMGIVETTKNPKNLGFLFWQGEAGEBYTGGDFIAAAVDGY	4093
Qy	593	LVRFLQSGEARLVED-PINDGWHRYTALRGRGSIQVDBEELVSGRS-PCPNVAV	640
Db	4094	VEYSMELDGEAVIRNSDRVNGERHVIAKKRDENATILEVD-RMLSHGTRPTSKKSM	4152
Qy	641	NAGSVYIGADPVALTGRSSGITGVKNLVLHSRPGAAPPQPLDQHRRAQAGANT	700
Db	4153	KLPGVAVGAGPLEVFTFRYKHNLNGCI--VVVEGETVG-----QINLSSAAVNGVNA	4205
Qy	701	RPCPS	705
Db	4206	NVCPA	4210

RESULT 9

Q81R8 Q81R8 PRELIMINARY; PRT; 4228 AA.

AC Q81R8; 01-MAR-2003 (T-EMBLrel. 23, Created)

DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE CG7981-PE.

GN TROL OR EG:BAC2593.11 OR CG7981.

OS Drosophila melanogaster (fruit fly).

OC Rukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227; [1]

RN SEQUENCE FROM N.A.

RP ADAMS M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Rukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yeaman C., Zeng Q., Chen L.X., Brandon R.C., Rogers Y.H., Blazej R.G., Champagne M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L., Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Balswin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Besu P.V., Berman B.P., Bhandari D., Bolshakov S., Borokova D., Botchan M.R., Bouck J., Brodeur G.M., Brotherton P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K.J., Evanselista C.C., Ferraz C., Ferriera S., Fleischmann W., Foeller C., Gabrielian A.C., Garg N.S., Gelbart W.M., Glasser K., Glisic S., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.D., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M., Palazon L.M., Patterson C., Pevzner S., Pohlman J., Pongor S., Poppe M., Reiter K., Remington K., Saunders R.D., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wasarman D.A., Weinstock G.M., Weissenbach J., Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhou X., Zhu S., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; "The genome sequence of Drosophila melanogaster.";

RT Science 287:2185-2195(2000).

RN SEQUENCE FROM N.A.

RP Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y., Banzon J., An H., Baldwin D., Banzone J., Beeson K.Y., Busam D.A., Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M., Dodson K.J., Dorssett V., Doup L.E., Doyle C., Dresnek D., Farfan D., Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A., Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J., Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A., McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J., Paclab J., Paragov S., Park S., Patel S., Pfeiffer B., Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F., Shouanavong S., Strong R., Swirskas R., Tector C., Tyler D., Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.; "Sequencing of Drosophila melanogaster genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases. [3]

RN SEQUENCE FROM N.A.

RP Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K., Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D., Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E., Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N., Krommiller B., Marshall B., Millburn G., Richter J., Russo S., Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.; "Annotation of Drosophila melanogaster genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases. [4]

RN SEQUENCE FROM N.A.

RP Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.; Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases. [5]

RN SEQUENCE FROM N.A.

RP FlyBase.

RL Submitted (SEP-2002) to the EMBL/GenBank/DBSJ databases.

RR EMBL; AB03424; AA09078.1;

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

Qy 333 ERLNKDGLRNGRPRVLRSSPGKSGQLNHLHTLLYLGVEPSPVLSLSPAINMSAHFRGCTVG 392
 Db 173 ERNGRKGALRVGDGPRVLRGSPVHTVNLKLEPLVYGGADPFKSLARAAAASVSSGFDGAIQ 232
 Qy 393 EYVNGKEL-----DLTVSFLSGIGQCYDSSPCER---OPCQHGATCMRAGEY 439
 Db 233 LVLSGGKQLLTPHVLKQVDVT-SFAG-----HPTCRASGHPCLNGASCVP---- 277
 Qy 440 EFOCLCRDGFKGDLCHEEENPCQLREPLRGHTCGQTRICLCLPFGSPRQQSGHGIAE 499
 Db 278 -----REAY-----VCLCPGFGSPHCKEGL----- 299
 Qy 500 SDWHELSGGNDARFGYCAFHDDGFLAFRCHVFSRSLPEVPEK-----TELEVRTS 551
 Db 300 -----VEKSAG-----DVTTLAFDGRITFVYLVNATVSEKALQNHLSIRTE 343
 Qy 552 TASGILLMQGVEVGEAGOGKDFISLGLQDGHVFRYQLGSGEARLVSSEDPINDGEWHRT 611
 Db 344 ATQGLAVLWS---SKATERADYVALAIVDGHLSYLNLSGQPVVLRSTVPTVNTNRMLRVV 399
 Qy 612 ALRERGRGSIQVDEELVSGSPGPNVANNKGSVYIGGAPD--VALTGGFRSSGHTGC 669
 Db 400 AHREQREGSLQVGNAPVPTGSSPLGATQDLDGALMLGGLPELFPVGPALPKAYGTGFPVGC 459
 Qy 670 VKNLVLSARPGAPPPDLDQHRQAQAGANTRPCPS 705
 Db 460 LRDVVVGR-----HPLHLEDAVTKPELRPCPT 487

RESULT 11
 Q9V714 ID Q9V714 PRELIMINARY; PRT; 1361 AA.
 AC Q9V714;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE CG8403 protein.
 GN SP2353 OR CG8403.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAINS=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yeung M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Durbin K.J., Evansgela C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrill J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J.R., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Ranson J.J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.B., Doyle C., Dresnek D., Farfan D.,
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Munoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "sequencing of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Seattle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RA EMBL; AE003808; AAF58071.2; -;
 DR HSP; P08709; 1BF9.
 DR FlyBase; FBgn0034070; SP2353.
 DR InterPro; IPR008985; ConA like_lec_gl.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00054; laminin_G; 3.
 DR SMART; SM00181; EGF; 4.
 DR SMART; SM00282; LAMG; 3.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 3.
 KW EGF-like domain.
 SQ SEQUENCE 1361 AA; 150298 MW; 4C71EDA81A12ABED CRC64;
 Qy Query Match 17.4%; Score 664.5; DB 5; Length 1361;
 Db Best Local Similarity 22.0%; Pred. No. 1e-40;
 Matches 221; Conservative 111; Mismatches 295; Indels 379; Gaps 23;
 Qy 1 EIKITFRPDSADGMLLYNGKRVFGSPTNLANRQDFISFLVGGPRFPRFDAGSGMATI 60
 Db 424 QLRTEFRPDSFDGILLISGER-----DRLTGDFMALLLNKGFVDFWFCDSGVS 474

QY 61 RHPTPLALGHFHTVTLRSLVGLDGLVAVNGTSGKFGQGLDNEELYLGGYDY-G 119
 Db 475 RSRETIILNENWVLIYRHRWDALVLMHGKTKVQGRSNGLSRITFRFVPLGGIGNITG 534
 QY 120 AIPKAGLSSGFGVRELIQGEIEIVFDHMLTAH-----GISHCPT--CRDRPCQ 168
 Db 535 LAKRPLAEGFAGCIR--RFVANE---HDYKTEHPLGDVINGFDIQDCSTDKCVRYPCQ 589
 QY 169 NGGQCHDSESSVYCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGVTCCRHL 228
 Db 590 HGGKCLPDDQGA-ICLCPIGFVGDLC----- 615
 QY 229 GRSGLRCBEGVTVTTSLGAGSYLALPALNTNTHLRLDVEPKPLAPDGVLLFSGKSG 288
 Db 616 ----IRMD----LQVPAFNGSSFLRYAPLGDLSALIWLKLVTLKPEQADGLIYSGPEHR 667
 QY 289 PVDFVSLAMVGGHLEFRYELGSLAVLSAEPLALGRWHRVSAERLNKDGSLRVNGGRP 348
 Db 668 --GDFIALYLNDDGVEFAFDLGGSPALVRSEHSLGOWHTIKISRTARLAVLKVXHQE 725
 QY 349 VLRSPPKSOGLNHLTLVYLGVEPSVPLSPATNMSAHFRGCVGEVSNKRLDLYSFL 408
 Db 726 VLTISNGFWHLSDQNLVFGVNVHVDRLFLDLKYPKFFVGCIDRINDINGSLGIVAEAL 785
 QY 409 GSQIGQCVDSSPCEROPCHGATCMPAGE-YEFQCLCRDGFKGDLCHEH--EENPCOLRE 465
 Db 786 GGSNIGNC--PHACVARPCGLAECVPQWESYE---CRCSIHNERCNKAAAEVPEQLPE 839
 QY 466 PCLHGTCQGT----- 477
 Db 840 LALHKSIVLETKDNGEAAKVKVSLGSLAHKSKHRNLHKPTATSTTTTSTTTTTEAP 899
 QY 478 ----- 481
 Db 900 SERTEATAGALSNEIEDDIIFRLVQQQQQKELKKHQOQTTPATSTTTSSGPKAK 959
 QY 482 PGFSG----- 486
 Db 960 PRLSGKHAASKEHHLKPNAAFTRKLSRLPHYESFQTNPDSDILTFEDNNDWVTSLQQ 1019
 QY 487 -----PRCQQ----- 492
 Db 1020 EYGDMAASQVPLAFEDASGTFRSDNNEEDENAFVDESIFDASDGTBEYQKLAQD 1079
 QY 493 -----SGHGAESDWHLEGS-----GNDAPQYQYAFHDDGFI----- 526
 Db 1080 MKRIMSNASHSHKKAQVQPPQSQSCEVTANEDTSQYSDYNDDELLTPVMQGBEVK 1139
 QY 527 -----APPG-----HV 532
 Db 1140 LEQHTSSTPQTHDWBLLKFDLSAQVQVGRKXNFACFAGSDSYFHYNDATMSQV 1199
 QY 533 FSRSLPEVETIEVLTASGLLMOGVEAGEQKDFISLQDGHVFRYQLGSG 592
 Db 1200 ISYSI-----DLNRIKTHSENGVILWTKRQ--GTTEHDDYLSLGTQGYLHFRYDLGSG 1253
 QY 593 EARL-VSEDPINDGEHRVTLAREGRGSIQVDGEELVSRGPNVAVNAGKSVYVIGGA 651
 Db 1254 EVDIREFNGTKVSDGLWHRVRAIRNSQEGYLEVDGRKVTIRAPKRLQNTWTGLYVGM 1313
 QY 652 PDVATLTGGRFSSGITGVKNVILHSGARPGAPPQPLDLQRAQAG 697
 Db 1314 PDVGYFTHQRYFSGIVGCISEIVLAGEMKLNFDNTLGTTEHNVETG 1359
 PRELIMINARY; PRT; 1361 AA.
 Q9NGV2
 AC Q9NGV2
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE SP2353. OR CG8403
 GN Drosophila melanogaster (Fruit fly).
 OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RA Serano T.L., Rendleton J.D., Rubin G.M.;
 RT "A reverse genetic screen for genes involved in Drosophila
 development.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF239610; AAF63502.1; -;
 DR HSSP; P08709; 1BF9.
 DR FlyBase; FBgn0034070; SP2353.
 DR InterPro; IPR008985; ConA_like_lec_gl.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002210; IEGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00054; laminin_G; 3.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00282; LamG; 3.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS50025; LAW_G_DOMAIN; 3.
 KW EGF-like domain.
 SQ SEQUENCE 1361 AA; 150355 MW; 206D9F9BF9EDFE47 CRC64;
 Query Match 17.4%; Score 664.5; DB 5; Length 1361;
 Best Local Similarity 22.0%; Pred. No. 1e-40;
 Matches 221; Conservative 111; Mismatches 295; Indels 379; Gaps 23;
 QY 1 EIKITRPSADQMLYNGKRVGSPFTMLNRQPDFISFGLVGGPFRFDAGSGMATI 60
 Db 424 QLRIEFRPSDFDGIILLSE-----DGLTGFALLLNKGVFEWDFGSGVGV 474
 QY 61 RHPTPLALGHFHTVTLRSLVGLDGLVAVNGTSGKFGQGLDNEELYLGGYDY-G 119
 Db 475 RSRETIILNENWVLIYRHRWDALVLMHGKTKVQGRSNGLSRITFRFVPLGGIGNITG 534
 QY 120 AIPKAGLSSGFGVRELIQGEIEIVFDHMLTAH-----GISHCPT--CRDRPCQ 168
 Db 535 LAKRPLAEGFAGCIR--RFVANE---HDYKTEHPLGDVINGFDIQDCSTDKCVRYPCQ 589
 QY 169 NGGQCHDSESSVYCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGVTCCRHL 228
 Db 590 HGGKCLPDDQGA-ICLCPIGFVGDLC----- 615
 QY 229 GRSGLRCBEGVTVTTSLGAGSYLALPALNTNTHLRLDVEPKPLAPDGVLLFSGKSG 288
 Db 616 ----IRMD----LQVPAFNGSSFLRYAPLGDLSALIWLKLVTLKPEQADGLIYSGPEHR 667
 QY 289 PVDFVSLAMVGGHLEFRYELGSLAVLSAEPLALGRWHRVSAERLNKDGSLRVNGGRP 348
 Db 668 --GDFIALYLNDDGVEFAFDLGGSPALVRSEHSLGOWHTIKISRTARLAVLKVXHQE 725
 QY 349 VLRSPPKSOGLNHLTLVYLGVEPSVPLSPATNMSAHFRGCVGEVSNKRLDLYSFL 408
 Db 726 VLTISNGFWHLSDQNLVFGVNVHVDRLFLDLKYPKFFVGCIDRINDINGSLGIVAEAL 785
 QY 409 GSQIGQCVDSSPCEROPCHGATCMPAGE-YEFQCLCRDGFKGDLCHEH--EENPCOLRE 465
 Db 786 GGSNIGNC--PHACVARPCGLAECVPQWESYE---CRCSIHNERCNKAAAEVPEQLPE 839
 QY 466 PCLHGTCQGT----- 477
 Db 840 LALHKSIVLETKDNGEAAKVKVSLGSLAHKSKHRNLHKPTATSTTTTSTTTTTEAP 899
 QY 478 ----- 481
 Db 900 SERTEATAGALSNEIEDDIIFRLVQQQQQKELKKHQOQTTPATSTTTSSGPKAK 959

QY 482 PGFSG----- 486
 Db 960 PRUSGRHASKHEHLKPNAAFRKLSRLPETHYESFQTFNPDSDILTFEDNNDWVTSLOQQ 1019
 QY 487 -----PRCQOQ----- 492
 Db 1020 EYGDAMAASQVPLAFEDASPGTPRSDNNEEDENAFVDESIFDASDGTFFYQRKQLAQD 1079
 QY 493 -----SGHGAESDWHLEGS-----GGNDAPGQYGFHDDGFL----- 526
 Db 1080 MKRIMSNAHSHKKAQVQFPQGSQEVGTANEDTSQYSDYDNDDELLTPVMQGSSEVK 1139
 QY 527 -----APFG-----HV 532
 Db 1140 LEQHTSTPQTHDMSLLKFLSAEHQSQVQVGRKNGFACFAGSDSYFHYNDADTMSQV 1199
 QY 533 FRSLSPEVETIELEVRTSASGLLWQVGEAGQKDFISLGLQDGHVFRYQLGSG 592
 Db 1200 ISYSI-----DNLRIKTHSENGVILWTGRQ-GTTEHDDYLSLGLTEQYVLFHFRYDLGSG 1253
 QY 593 EARL-VSEDPINDEWHRYTALREGRGSIQVYDGBELVSRGPFENAVNAKGSVYICGA 651
 Db 1254 EVDIREFNGTKVSDGLWHRVRAIRNSQGVLEVDGRKVTILRAPKLRQLRQNTDTGLVYGGM 1313
 QY 652 PDVATLTGGFRSGITGVNVLVHSARPGAPPDPLDLQHRQAAG 697
 Db 1314 PDVGYFTHQYFSGIVGICISEIVLAGEMKLNFPNTLGTENHVEITG 1359

RESULT 13
 Q8NAL2 PRELIMINARY; PRT; 463 AA.
 AC Q8NAL2: 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical protein FLJ35160.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
 RA Otaki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
 RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
 RA Muragano S., Nagahari K., Masuno Y., Nagai K., Isogai T.;
 RA "NEDO human cDNA sequencing project."
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AK092479; BAC03900.1;
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00054; Laminin_G; 2.
 DR SMART; SM00181; EGF_2.
 DR SMART; SM00179; EGF_Ca; 2.
 DR SMART; SM00282; Laug; 2.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS50025; LAM_G DOMAIN; 2.
 KW Hypothetical protein; EGF-like domain.
 SQ SEQUENCE 463 AA; 50635 MW; F6889F2714D5D0EC CRC64;

Query Match 16.4%; Score 628; DB 4; Length 463;
 Best Local Similarity 28.7%; Pred. No. 1.3e-38;
 Matches 162; Conservative 80; Mismatches 196; Indels 124; Gaps 16;

QY 151 LTAHGISHCPT--CRDRPCQGGCHDSSESSVYCVCPAGFTGSRCEHSQALHCHPEACG 208
 Db 10 LSGADVGECSGICDEASCIHGGTCTAKADSVICLCPGLG----- 50
 QY 209 PDATCVNRPDGRGYTCRHLGRSGLRCBEGVITVTFSL--SGAGSYLALPALPALTHTHEL-- 265
 Db 51 -----KGRHCEDAFTLTIQFRESLSRYAATPWFLEPQHYLSF 88
 QY 266 -RLDVEKPLADPQVLLFS---GKSGFVDFVLSLWVGCHLEFPFRELGSGLAVLRSAEP 321
 Db 89 MEFEITFRPDSGDVLLSYDYDTGSK-----DFLSINLAGGHVEFRFCGSGTGVLREDDP 143
 QY 322 LALGRWHRVSAERLNKDGSLRVNGRPLVLRSSPKSQSLNLHLTLILYLGVEPSPVPLSPAT 381
 Db 144 LTLGNWHELVRSTAKNGILQVDKQKIVEGMAEGGFTQIKCNTDIFIGGVFNVDVVKNS 203
 QY 382 NMSAHFGCVGEVSVNGKRLDLTVSFLGSGOIGOCYDSSPCERQPCOHGATCMPAGEYEF 441
 Db 204 GVLKPFPSGSIQKILNDRTIHVRHDF--TSGVNVENAAHFVCRAPCAHGGSCRPKE--GY 260
 QY 442 QCLCRDGFKGLCBEHENPCQLREPCPLHGGTCCQT--RCLCLPGFSGPRCQOQSGHGIAE 499
 Db 261 DCDPLGPEG-----LH--CQKAIIEAIEIFQFIG----- 288
 QY 500 SDWHEGSGNDAPGQYGFHDDGFLAFPGHFVRSRSLPEVETIELEVRTSASGLLW 559
 Db 289 -----RSYLTYDNDPDLKRVSG---SRS-----NVFMRFKTKAKDGLLW 325
 QY 560 QGVEVEAGQKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLRGRRG 619
 Db 326 RG--DSPMRFNSDFISLGRDGLVFSYNLGSGVASIWNVNGSFNDGRWHRVKAVRDQSG 383
 QY 620 SIQVDEELVSGRSPFNVAVNAKGSVYIGAPDVAITLGTGRFSSGITGVNVLVHSAR 679
 Db 384 KITVDYDYGARTGKSPGMMRQLNGLYVGMKBIALHTRQYMRGLVGCISHFTLST-- 441
 QY 680 FGAPPPQLDLQHRQAAGNTRPC 703
 Db 442 -----DYHISLVEADVQGNINTC 460

RESULT 14
 Q8BU56 PRELIMINARY; PRT; 822 AA.
 AC Q8BU56:
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Hypothetical EGF-like domain.
 GN AU040377.
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Oviduct;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The PANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK087659; BAC39956.1; -
 DR MGD; MGI:2146149; AU040377.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR003962; FNIII_subd.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR008957; FN_III-like.

```

DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00054; laminin_G; 2.
DR PRINTS; PR00014; FNVVPEIII.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00282; LamG; 2.
DR PROSITE; PS00022; EGF 1; 2.
DR PROSITE; PS01186; EGF 2; 1.
DR PROSITE; PS50025; Lam_G_DOMAIN; 2.
KW Hypothetical protein.
SQ SEQUENCE 822 AA; 89669 MW; 225A01D56D8BE691 CRC64;

Query Match 15.2%; Score 581.5; DB 11; Length 822;
Best Local Similarity 30.3%; Pred. No. 8.7e-35;
Matches 139; Conservative 70; Mismatches 184; Indels 65; Gaps 10;

QY 200 LHCHEACGPDATCVNRDPDGGYTCRCHLGRSLRCRBEVTVVTPSLGAGSYLALPALT 259
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 345 MSCDETLCSAFCVNDYANGSRCHCNLGGKGEACSEDFIQYQPFPG-HSYVTFPLX 403
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 NTHHELRLDVEFKPLAPPVLLFSGKSPVEDFVSLAMVGGHLEFRYELGSLAVLRSR 319
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 404 NSYQAFQVLEPRAEADGLLLYCGESEHGRGDFMSLALIRSLHFRFNCGTGIAIISE 463
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 320 EPLALGRWHRYSAERLNKDGSLRVNGRPLVRSRSGKSGQLNHTLTYLGGVPSVPLSP 379
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 464 TKIKLGARWHVTVLYRDGLNGLMLQNLNNGIPVTGQSQGQSKIFRTPLYLGGAPRAYLVR 523
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 380 ATNMSAHRFGCYEVSVNGKRLDLYFLG----SQGTCYQVSSPCERQPCQHGATCMP 435
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 524 ATGTRNGFQGCQSLSVNGKIDMRPWLGRALANGADVGECSGICDEASCIHGTCAA 582
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 436 AGEYFQCLCRDQFQDLCEHENFCQLREPLHGCTCGTRCLCLPGFSPRCQGSQH 495
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 583 IKADSYICLCPGFRG---RH-----CEDAFALTIQF-----RESLR 617
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 496 GIAESDWHLEGSGGNDAPQYGAYPHDDGLAFFPGHFRSRLPEVPTIELEVRTSTASG 555
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 618 SYAATPWPLE-----PQHLSFT-----EFEITFRPDSGDG 648
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 556 LLLMGQVVEAGEQKDFISLQDHLVFRVLQSGEARLVSEDPINDGEMHWRTALRE 615
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 649 VLLY-SYDTG-----SKDFLSINVAAGHVEFRFDCGGTGLRSEAPLTGQWHLDRVSR 703
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 616 GRGSIQVDBELVSRSPGNVAVNAKGVYIGAPD 653
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 704 AKNGILQVDKQKVEGMAEGGFTQIKCNTDIFIGVFN 741
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 15
QYQV01 PRELIMINARY; PRT; 1039 AA.
AC QV031;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE CG7245 protein.
GN CG7245.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

```

```

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davey P., Davis L., Dew I., Dietz S.M.,
RA de Fabios B., Delcher A., Deng Z., Mays A.D., Mays R.C., Dunn P.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Limmi B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Morkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelzson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidgen-Kiamos I., Simpson M., Skupeki M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:12185-2195(2000).
DR EMBL; AE003583; AAF51266.1; -.
DR HSP; P00740; LEDM.
DR FlyBase; FBgn0031416; CG7245.
DR GO; GO:0003676; Funnucleic acid binding; IEA.
DR InterPro; IPR008985; ConA_like_lec_g1.
DR InterPro; IPR000742; EGF-2.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR001791; Laminin G.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF00034; laminin_G; 3.
DR SMART; SM00181; EGF; 6.
DR SMART; SM00282; LamG; 3.
DR PROSITE; PS00022; EGF 1; 6.
DR PROSITE; PS01186; EGF 2; 3.
DR PROSITE; PS50025; Lam_G_DOMAIN; 4.
DR PROSITE; PS00030; REM_RNP_1; 1.
DR EGF-like domain.
KW SEQUENCE 1039 AA; 114305 MW; 8864E2A0F345590 CRC64;

Query Match 15.1%; Score 576.5; DB 5; Length 1039;
Best Local Similarity 27.2%; Pred. No. 2.8e-34;
Matches 200; Conservative 95; Mismatches 298; Indels 143; Gaps 31;

QY 70 HFHTVIL--LRSLT--QGSLVGDLPVNGTSQKFGQLD-----INEELYLGGYV 116
Db 2 HCSTIRLDFRNRYTHCNASLLVNDTLAMSGDQFTWKLKLPRLHTPEALNTWLLGGAP 61
QY 117 D--YGAI---PKAGLSSGFTGCVRELRIQG-----EIVVFHDLN-----LTAHGI 156
Db 62 QAPIGLIIELPPAQSQSGSFTGLTLRLNGQAREIFQPDKDKISKELGSRSDALDFGI 121
QY 157 SHCPT--CRDRPCQNGQC-----HDSSESSVVCVPAGFTGSRCEHSQLHCHP 204
Db 122 TEGSLACLSSPCRNGAACIKIETNLDENGKAEKWKCKCPTGYMPTCEISV---CED 178

```

```

Qy 205 EACGPDATCVNRPDGRGYTCRCHLGRSGLRCEBEGVTVTTPSLSGA-----GSYLALPALTN 260
Db 179 NPCOYGGTCVQFP--GSGYLCLCEPLGKHGHYCEHLEVALFSPFGSVNGLSSFVAYTVPIP 237
Qy 261 THEHLRLDVEFKP--LAPDGVLLFSG--GKSGPVEDFVSLAMVGGHLEFRYELGSLAYL 316
Db 238 LEYSLELSFKILPOTMSQISLLAFFGSGYHDEKSDHLAVSFIQGYIMLTWNLGAGPRI 297
Qy 317 RSAPPL-----ALGRWHRVSAERLNKDGSLRVNGRFPVLRSPGKSOGLNHLHLLYLGGV 371
Db 298 FTOKPIDFRDLAPRVPEIKVGRIGRQAWLSVDGKFNITGRSPGSGRMDVLPILYLGGH 357
Qy 372 E-----FSVPLSPATMSAHPFGCVGVSVNGKR-----LDLTSFGLSQGIGOCYDS 419
Db 358 EIANFNTLPHDLPL-----HSGFQGCIIYDVLKAGQVTVPLQETRCVRG--RGVQC--GT 409
Qy 420 SPCERQPCQHGATCMPAGEYEFQCLCRDPFKGDLCEHEENPC--OLREPCLHGTC-----Q 474
Db 410 RECHRAHQHDGACLQHGA--TFTCICQEGWYGPLCAOPTNPCDSFNKCYEDATCVPLVN 468
Qy 475 GTRCLCLPFGSPRCQOQSGHGIAESDWHLEGS-----GGNDAPG-----QYGA 518
Db 469 GYECDCPVGRTGKNCCEVIR---SLSVSLTGRRSYLAVRWPYLYDGGDKLGAQRSQMS 525
Qy 519 YFHDDGFLAFP-----GHVFSRSLPEVPET-----IELE 547
Db 526 YRNFTKLLMPKPIPTPSSHVFMKLLNEVEKORSFSPVPLMGSKSFEHHRVQFFFIEFQ 585
Qy 548 VRTSTAGLLLMQVEVGEAGQKDFISLGLDGHLYFRYQLGSGEARLY--SEDPINDGE 606
Db 586 LRPLSERGLLYFGLNNDKIKIGFVSLSLQGGVVEFRISGPNHVTVRSVRLAIGE 645
Qy 607 WHRVTLRGRGSIQVDGEBELVSGRSFGRPNVAVNAKSVVYGGAPDVAITLGGFRSS-- 664
Db 646 WHKIMQAQRGWLTLWVFG--SASSALAFSAEVLVEPDSLLYIGGLKDVSKLPHNAISGFP 704
Qy 665 -GITCCVKNLVLSAR 679
Db 705 IPFRGCVRGLVVSQTR 720

```

Search completed: March 9, 2004, 17:21:42
Job time : 67.739 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: March 9, 2004, 17:19:48 ; Search time 85.8465 Seconds
(without alignments)
2320.373 Million cell updates/sec

Title: US-10-006-011A-3
Perfect score: 705
Sequence: 1 EKITFRPDSADGMLLNGQ.....QLDLQHRAGAGANTRPCPS 705

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : A_Geneseq_29Jan04.*
- 1: Geneseqp1980s.*
 - 2: Geneseqp1990s.*
 - 3: Geneseqp2000s.*
 - 4: Geneseqp2001s.*
 - 5: Geneseqp2002s.*
 - 6: Geneseqp2003as.*
 - 7: Geneseqp2003bs.*
 - 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	705	100.0	4391	6	AAE34390	Aae34390 Human per
2	443	62.8	4393	4	AAB31889	Aab31889 Amino aci
3	383	54.3	4436	4	ARG23265	Arg23265 Novel hum
4	195	27.7	195	4	AAB31890	Aab31890 Amino aci
5	18	2.6	18	4	ABB55911	Abb55911 Vascular
6	14	2.0	14	4	ABB55908	Abb55908 Vascular
7	14	2.0	14	4	ABB55905	Abb55905 Vascular
8	10	1.4	10	4	ABB55910	Abb55910 Vascular
9	9	1.3	15	4	ABB55909	Abb55909 Vascular
10	9	1.3	159	6	ABM65732	Abm65732 Propionib
11	9	1.3	2531	7	AD663713	Ad663713 Rat Prote
12	9	1.3	2531	7	AD663705	Ad663705 Rat Prote
13	9	1.3	2531	7	AD663709	Ad663709 Rat Prote
14	9	1.3	2531	7	AD663701	Ad663701 Rat Prote
15	9	1.3	3319	4	ABB70376	Abb70376 Drosophi
16	8	1.1	16	4	ABB55907	Abb55907 Vascular
17	8	1.1	14	4	ABB55906	Abb55906 Vascular
18	8	1.1	36	5	AAE13652	Aae13652 Mouse cyc
19	8	1.1	47	5	AAE13653	Aae13653 Human e-s
20	8	1.1	54	4	AAU40544	Aau40544 Propionib
21	8	1.1	54	6	ABM37063	Abm37063 Propionib
22	8	1.1	54	7	ABO23342	Abo23342 Human sec
23	8	1.1	55	2	AAU01436	Aau01436 Secreted
24	8	1.1	66	5	AAE13647	Aae13647 Human fib
25	8	1.1	77	4	AAU63351	Aau63351 Propionib

ID	AAE34390 standard; protein; 4391 AA.	XX	AC	AAE34390;	XX	14-MAY-2003 (first entry)	XX	Human perlecan protein.	XX	Human; diagnosis; osteoarthritis; rheumatoid arthritis; perlecan.	OS	Homo sapiens.	XX	WO200295415-A2.	XX	28-NOV-2002.	XX	22-MAY-2002; 2002WO-EP005612.	XX	23-MAY-2001; 2001GB-00012626.	XX	(OSTE-) OSTEOMETER BIO TECH AS.	XX	Christgau S, Henriksen DB, Cloos PAC;	XX	WPI; 2003-140389/13.	XX	An assay for the diagnosis or assessment of the severity of osteoarthritis or rheumatoid arthritis comprising detecting an isomerized or optically inverted protein in a sample.	XX	Disclosure; Page 46-67; 106pp; English.	XX	The invention relates to an assay for the diagnosis or assessment of the severity of osteoarthritis or rheumatoid arthritis. The assay involves measuring (in a biological sample) the amount or presence of an isomerized or optically inverted protein or one or more isomerized or optically inverted fragments from proteins such as perlecan, Biglycan, decorin, fibrillin-1 or protocadherin. The assay is useful for the diagnosis or assessment of the severity of osteoarthritis or rheumatoid arthritis. The present sequence is human perlecan protein	XX	Sequence 4391 AA;	XX	Query Match 100.0%; Score 705; DB 6; Length 4391;	XX	Best Local Similarity 100.0%; Pred. NO. 0;	XX	Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
26	8	1.1	77	6	ABMS9870	Abms9870 Propionib																																		
27	8	1.1	80	4	AAE15081	Aae15081 Peptide #																																		
28	8	1.1	80	4	ABB35070	Abb35070 Peptide #																																		
29	8	1.1	80	4	AAE28574	Aae28574 Peptide #																																		
30	8	1.1	80	4	ABB29890	Abb29890 Peptide #																																		
31	8	1.1	80	4	ABB20486	Abb20486 Protein #																																		
32	8	1.1	80	4	AAE68258	Aae68258 Human bon																																		
33	8	1.1	80	4	AAE55885	Aae55885 Human bra																																		
34	8	1.1	80	4	ABC49912	Abc49912 Human liv																																		
35	8	1.1	80	4	AAE03808	Aae03808 Peptide #																																		
36	8	1.1	80	5	ABG37793	Abg37793 Human pep																																		
37	8	1.1	82	5	AAE13646	Aae13646 Human coa																																		
38	8	1.1	83	5	AAE13645	Aae13645 Pig facto																																		
39	8	1.1	88	5	AAE13648	Aae13648 Human coa																																		
40	8	1.1	110	5	ABBS3124	Abbs3124 Human ORF																																		
41	8	1.1	123	5	AAE13650	Aae13650 Human ins																																		
42	8	1.1	131	5	ABB90171	Abb90171 Human pol																																		
43	8	1.1	143	5	AAE13643	Aae13643 Mouse lam																																		
44	8	1.1	144	5	ABBS3123	Abbs3123 Human ORF																																		
45	8	1.1	144	5	ABBS3141	Abbs3141 Human ORF																																		

ALIGNMENTS

QY 1 EIKITRPPDSADGMLLYNGQKRVPSGPTNLANKRQDFISFGLVGGREPERFRDAGSGWATI 60
 Db 3687 EIKITRPPDSADGMLLYNGQKRVPSGPTNLANKRQDFISFGLVGGREPERFRDAGSGWATI 3746
 QY 61 RHPTPLALGHFHTVTLRSLTQGLSLVGLDAPVNGFTSQKFGQLDNEELYLGGYPDYGA 120
 Db 3747 RHPTPLALGHFHTVTLRSLTQGLSLVGLDAPVNGFTSQKFGQLDNEELYLGGYPDYGA 3806
 QY 121 IPKAGLSGSGIGCVRELRIGQEIIVFDLNLTAHGISHCPTCRDRPCONGQCHDSSESS 180
 Db 3807 IPKAGLSGSGIGCVRELRIGQEIIVFDLNLTAHGISHCPTCRDRPCONGQCHDSSESS 3866
 QY 161 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGHTCRCHLGRSGLRCBEGVT 240
 Db 3867 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGHTCRCHLGRSGLRCBEGVT 3926
 QY 241 VTTPSLGGAGSYLALPALNTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
 Db 3927 VTTPSLGGAGSYLALPALNTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 3986
 QY 301 GHLEFRYELGSLAVLRSAPLALGRWHRVSAERLNKDGSLRVNGRVLRSFPGKSQGL 360
 Db 3987 GHLEFRYELGSLAVLRSAPLALGRWHRVSAERLNKDGSLRVNGRVLRSFPGKSQGL 4046
 QY 361 NLHTLLYLGVEVPSVPLSPATNMSAHRFCVGEVSVNGKRLDLYSFLGSGQIGQCYDSS 420
 Db 4047 NLHTLLYLGVEVPSVPLSPATNMSAHRFCVGEVSVNGKRLDLYSFLGSGQIGQCYDSS 4106
 QY 421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGLDCEHEENPCQLRPECLHGTCQGRTRCLC 480
 Db 4107 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGLDCEHEENPCQLRPECLHGTCQGRTRCLC 4166
 QY 481 LPFGSGPRCOQSGHGIASDWHLEGGGNDAPGQYGFYHDDGFLAFPGHVSRSLEPEV 540
 Db 4167 LPFGSGPRCOQSGHGIASDWHLEGGGNDAPGQYGFYHDDGFLAFPGHVSRSLEPEV 4226
 QY 541 PETIELEVRSTASGLLWQGVVEGAGQKDFISLGLDQHLVFRYQLGSGEARLYSEED 600
 Db 4227 PETIELEVRSTASGLLWQGVVEGAGQKDFISLGLDQHLVFRYQLGSGEARLYSEED 4286
 QY 601 PINDGEHRTALRREGRRGSIQVDGELYSGRSPGNVAVNAKGSVYIGGAPDVATLTGG 660
 Db 4287 PINDGEHRTALRREGRRGSIQVDGELYSGRSPGNVAVNAKGSVYIGGAPDVATLTGG 4346
 QY 661 RFSGGITGVKLVLSARPGAPPDLOHRAQAGANTRECPS 705
 Db 4347 RFSGGITGVKLVLSARPGAPPDLOHRAQAGANTRECPS 4391

RESULT 2
 AAB31889
 ID AAB31889 standard; protein; 4393 AA.
 AC AAB31889;
 XX
 XX 15-MAY-2001 (first entry)
 DT
 XX Amino acid sequence of a human protein.
 DE Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX
 OS Homo sapiens.
 XX
 XX WO200105422-A2.
 FN
 XX 25-JAN-2001.
 PD
 XX 17-JUL-2000; 2000WO-FR002057.
 PF

XX
 PR
 XX
 PA (INMR) BIOMERIEUX STELHYS.
 XX
 PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX WPI; 2001-159475/16.
 XX
 PT Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 DR polypeptides or related nucleic acid or ligand.
 XX
 PS Claim 1; Page 138-152; 209pp; French.
 XX
 CC The present sequence represents a human protein, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 4393 AA;

Query Match 62.8%; Score 443; DB 4; Length 4393;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 643; Conservative 2; Mismatches 0; Gaps 0;
 QY 1 EIKITRPPDSADGMLLYNGQKRVPSGPTNLANKRQDFISFGLVGGREPERFRDAGSGWATI 60
 Db 3689 EIKITRPPDSADGMLLYNGQKRVPSGPTNLANKRQDFISFGLVGGREPERFRDAGSGWATI 3748
 QY 61 RHPTPLALGHFHTVTLRSLTQGLSLVGLDAPVNGFTSQKFGQLDNEELYLGGYPDYGA 120
 Db 3749 RHPTPLALGHFHTVTLRSLTQGLSLVGLDAPVNGFTSQKFGQLDNEELYLGGYPDYGA 3808
 QY 121 IPKAGLSGSGIGCVRELRIGQEIIVFDLNLTAHGISHCPTCRDRPCONGQCHDSSESS 180
 Db 3809 IPKAGLSGSGIGCVRELRIGQEIIVFDLNLTAHGISHCPTCRDRPCONGQCHDSSESS 3868
 QY 161 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGHTCRCHLGRSGLRCBEGVT 240
 Db 3869 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGHTCRCHLGRSGLRCBEGVT 3928
 QY 241 VTTPSLGGAGSYLALPALNTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
 Db 3929 VTTPSLGGAGSYLALPALNTNTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVG 3988
 QY 301 GHLEFRYELGSLAVLRSAPLALGRWHRVSAERLNKDGSLRVNGRVLRSFPGKSQGL 360
 Db 3989 GHLEFRYELGSLAVLRSAPLALGRWHRVSAERLNKDGSLRVNGRVLRSFPGKSQGL 4048
 QY 361 NLHTLLYLGVEVPSVPLSPATNMSAHRFCVGEVSVNGKRLDLYSFLGSGQIGQCYDSS 420
 Db 4049 NLHTLLYLGVEVPSVPLSPATNMSAHRFCVGEVSVNGKRLDLYSFLGSGQIGQCYDSS 4108
 QY 421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGLDCEHEENPCQLRPECLHGTCQGRTRCLC 480
 Db 4109 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGLDCEHEENPCQLRPECLHGTCQGRTRCLC 4168
 QY 481 LPFGSGPRCOQSGHGIASDWHLEGGGNDAPGQYGFYHDDGFLAFPGHVSRSLEPEV 540
 Db 4169 LPFGSGPRCOQSGHGIASDWHLEGGGNDAPGQYGFYHDDGFLAFPGHVSRSLEPEV 4228

QY 541 PETIELEVRTSTASGLLWQVGEAGQKDFISLGLQDHLVFRYQLGSGEARLYSED 600
 |||||
 Db 4229 PETIELEVRTSTASGLLWQVGEAGQKDFISLGLQDHLVFRYQLGSGEARLYSED 4288
 |||||
 QY 601 PINDGEHWRVTRALRGRRGSIQVDGEELVSGRSPGNVAVNAKGS 645
 |||||
 Db 4289 PINDGEHWRVTRALRGRRGSIQVDGEELVSGRSPGNVAVNAKGS 4333
 |||||

RESULT 3
 ABG23265
 ID ABG23265 standard; protein; 4436 AA.
 XX

AC ABG23265;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #23256.
 XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX

OS Homo sapiens.
 XX
 PN WC200175067-A2.
 XX
 XX 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US008631.
 XX
 XX 31-MAR-2000; 2000US-00540217.
 PR
 XX 23-AUG-2000; 2000US-00649167.
 PR
 XX (HYSE-) HYSEQ INC.
 PA

XX Drmanac RT, Liu C, Tang YT;
 XX
 XX WPI; 2001-639362/73.
 DR
 XX N-PSDB; AAS87452.
 DR

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX

PS Claim 20; SEQ ID NO 53624; 103pp; English.
 XX

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 4436 AA;

Query Match 54.3%; Score 383; DB 4; Length 4436;
 Best Local Similarity 99.7%; Pred. NO. 0;

Matches 583; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 EIKITFRPDSADQMLLYNGQKRVPGSFTNLANRPDPFISFGLVGGRRPFRFDAGSGMATI 60
 |||||
 Db 3722 EIKITFRPDSADQMLLYNGQKRVPGSFTNLANRPDPFISFGLVGGRRPFRFDAGSGMATI 3781
 |||||
 QY 61 RHPTPLALGHFHTVLLRSITQGSLLVGDLPVNGTSQKFGQLDNEELYLGGYDYGA 120
 |||||
 Db 3782 RHPTPLALGHFHTVLLRSITQGSLLVGDLPVNGTSQKFGQLDNEELYLGGYDYGA 3841
 |||||
 QY 121 IPKAGLSSGFIGCVRELRIQGEBIVFHDNLMLTAHGISHCPTCRDRPCQNGGQCHDSSESS 180
 |||||
 Db 3842 IPKAGLSSGFIGCVRELRIQGEBIVFHDNLMLTAHGISHCPTCRDRPCQNGGQCHDSSESS 3901
 |||||
 QY 181 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 240
 |||||
 Db 3902 YVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGLRCEEGVT 3961
 |||||
 QY 241 VVTPSLGAGSYLALPALTNTHHELRLDVEFKELAPDGVLLFSGGKSGPVYDFVSLAMVG 300
 |||||
 Db 3962 VVTPSLGAGSYLALPALTNTHHELRLDVEFKELAPDGVLLFSGGKSGPVYDFVSLAMVG 4021
 |||||
 QY 301 CHLEFRYELGSLAVLRSAPFLALGRHRYSAERLNKDSLRVNGGEPVLRSSPGKSQGL 360
 |||||
 Db 4022 CHLEFRYELGSLAVLRSAPFLALGRHRYSAERLNKDSLRVNGGEPVLRSSPGKSQGL 4081
 |||||
 QY 361 NLHTLLYLGVEPSVPLSPATNMSAHPRCVGEVSVNGKRLDLTYSFLGSGGICQCYDSS 420
 |||||
 Db 4082 NLHTLLYLGVEPSVPLSPATNMSAHPRCVGEVSVNGKRLDLTYSFLGSGGICQCYDSS 4141
 |||||
 QY 421 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCHEEENPCOLREPCLRHGTCCGTCLC 480
 |||||
 Db 4142 PCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCHEEENPCOLREPCLRHGTCCGTCLC 4201
 |||||
 QY 481 LPFGSPRCQOQSGHGIAESDWHLEGGSGNDAPQYGYAFHDDGFLAAPPGHVFRSLPEV 540
 |||||
 Db 4202 LPFGSPRCQOQSGHGIAESDWHLEGGSGNDAPQYGYAFHDDGFLAAPPGHVFRSLPEV 4261
 |||||
 QY 541 PETIELEVRTSTASGLLWQVGEAGQKDFISLGLQDHLV 585
 |||||
 Db 4262 PETIELEVRTSTASGLLWQVGEAGQKDFISLGLQDHLV 4306
 |||||

RESULT 4
 AAB31890
 ID AAB31890 standard; protein; 195 AA.
 XX

AC AAB31890;
 XX
 DT 15-MAY-2001 (first entry)
 XX

DE Amino acid sequence of the C-terminal of the human perlecan protein.
 XX

KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
 XX

OS Homo sapiens.
 XX

PN WO200105422-A2.
 XX

PD 25-JAN-2001.
 XX

XX 17-JUL-2000; 2000WO-FR002057.
 XX

PR 15-JUL-1999; 99FR-00009372.
 XX

PA (INMR) BIOMERIEUX STELHYS.
 XX

PI Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
 XX

DR WPI; 2001-159475/16.
 DR N-PSDB; AAF54728.
 XX
 XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 XX Claim 1; Page 152-153; 209pp; French.
 XX
 CC The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC perlecan, precursor of the retinol-binding plasma protein precursor of
 CC the ganglioside GM2 activator, calgranulin B or sapsin B protein
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 XX Sequence 195 AA;

Query Match 27.7%; Score 195; DB 4; Length 195;
 Best Local Similarity 100.0%; Pred. No. 5.3e-183; Indels 0; Gaps 0;
 Matches 195; Conservative 0; Mismatches 0;
 QY 511 DAPGOYGFYFDDGFLAPFGVFRSLEPEVPTIELEVRTSTASGLLWQGVGEAGQG 570
 DB 1 DAPGOYGFYFDDGFLAPFGVFRSLEPEVPTIELEVRTSTASGLLWQGVGEAGQG 60
 QY 571 KDFISLQDGHVFRYQLSGEARLVSEDPINDGEWHRVTLREGRGSTOVDGEEELVS 630
 DB 61 KDFISLQDGHVFRYQLSGEARLVSEDPINDGEWHRVTLREGRRGSLQVDGEEELVS 120
 QY 631 GRSPGNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKLVLHSGARPGAPPPLDL 690
 DB 121 GRSPGNVAVNAKGSVYIGGAPDVATLTGGRFSSGITGCVKLVLHSGARPGAPPPLDL 180
 QY 691 QHRAQAGANTRPCPS 705
 DB 181 QHRAQAGANTRPCPS 195

RESULT 5
 ABB55911
 ID ABB55911 standard; peptide; 18 AA.
 AC ABB55911;
 XX
 XX 15-FEB-2002 (first entry)
 DT
 XX Vascular dementia-associated protein isoform (VPI) 111.
 DE
 XX Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
 KW diagnosis; prognosis; gene therapy.
 XX
 XX Homo sapiens.
 OS
 XX WO200169261-A2.
 PN
 XX 20-SEP-2001.
 PD
 XX 14-MAR-2001; 2001WO-GB001106.
 PF
 XX 15-MAR-2000; 2000GB-00006285.
 PR 24-NOV-2000; 2000GB-00028734.
 PR 28-NOV-2000; 2000US-00724391.
 XX

(OXFO-) OXFORD GLYCOSCIENCES UK LTD.
 Herath HMAC, Parekh RB, Rohlf C;
 WPI; 2001-557937/62.
 Screening, diagnosis or prognosis of vascular dementia (VD), useful for
 determining stage of VD and monitoring the effect of VD therapy,
 comprises analyzing body fluid by 2-dimensional electrophoresis for
 features correlated with VD.
 Claim 6; Page 32; 151pp; English.
 The invention relates to screening, diagnosis or prognosis of Vascular
 Dementia (VD) in a subject comprising analysing body fluid from the
 subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
 features containing at least one chosen feature whose relative abundance
 correlates with the presence, absence, stage or severity of VD or
 predicts the onset or course of VD, especially detecting in a sample of
 cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
 protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
 specification. Detecting VD-associated features and VPI is useful for the
 screening, diagnosis or prognosis of VD, for determining the stage or
 severity of VD, for identifying a subject at risk of VD or for monitoring
 the effect of therapy administered to a subject having VD. Nucleic acids
 encoding a VPI or inhibiting the function of a VPI are useful for the
 treatment of VD and for gene therapy
 XX
 XX Sequence 18 AA;

Query Match 2.6%; Score 18; DB 4; Length 18;
 Best Local Similarity 100.0%; Pred. No. 9.5e-10;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 644 GSVYIGGAPDVATLTGGR 661
 DB 1 GSVYIGGAPDVATLTGGR 18

RESULT 6
 ABB55908
 ID ABB55908 standard; peptide; 14 AA.
 AC ABB55908;
 XX
 XX 15-FEB-2002 (first entry)
 DT
 XX Vascular dementia-associated protein isoform (VPI) 108.
 DE
 XX Vascular Dementia; VD; VD-associated protein isoform; VPI; screening;
 KW diagnosis; prognosis; gene therapy.
 XX
 XX Homo sapiens.
 OS
 XX WO200169261-A2.
 PN
 XX 20-SEP-2001.
 PD
 XX 14-MAR-2001; 2001WO-GB001106.
 PF
 XX 15-MAR-2000; 2000GB-00006285.
 PR 24-NOV-2000; 2000GB-00028734.
 PR 28-NOV-2000; 2000US-00724391.
 XX
 XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
 Herath HMAC, Parekh RB, Rohlf C;
 WPI; 2001-557937/62.
 Screening, diagnosis or prognosis of vascular dementia (VD), useful for
 determining stage of VD and monitoring the effect of VD therapy,
 comprises analyzing body fluid by 2-dimensional electrophoresis for
 PT

PT features correlated with VD.

XX Claim 6; Page 32; 151pp; English.

CC The invention relates to screening, diagnosis or prognosis of Vascular Dementia (VD) in a subject comprising analysing body fluid from the subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of features containing at least one chosen feature whose relative abundance correlates with the presence, absence, stage or severity of VD or predicts the onset or course of VD, especially detecting in a sample of cerebrospinal fluid (CSF) from the subject one of 223 VD-associated protein isoforms (VPIs) (ABB5801-ABB56295) as fully defined in the specification. Detecting VD-associated features and VPI is useful for the screening, diagnosis or prognosis of VD, for determining the stage or severity of VD, for identifying a subject at risk of VD or for monitoring the effect of therapy administered to a subject having VD. Nucleic acids encoding a VPI or inhibiting the function of a VPI are useful for the treatment of VD and for gene therapy

XX SQ Sequence 14 AA;

Query Match 2.0%; Score 14; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 6.5e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 596 LVSEDPINDGEVHR 609
|||||
1 LVSEDPINDGEVHR 14

RESULT 7

ABB5905

ID ABB55905 standard; peptide; 14 AA.

XX AC ABB55905;

XX DT 15-FEB-2002 (first entry)

XX DE Vascular dementia-associated protein isoform (VPI) 105.

XX KW Vascular Dementia; VD; VD-associated protein isoform; VPI; screening; diagnosis; prognosis; gene therapy.

XX OS Homo sapiens.

XX PN W0200169261-A2.

XX PD 20-SEP-2001.

XX PF 14-MAR-2001; 2001WO-GB001106.

XX PR 15-MAR-2000; 2000GB-00006285.

XX PR 24-NOV-2000; 2000GB-00028734.

XX PR 28-NOV-2000; 2000US-00724391.

XX PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.

XX PI Herath HWAC, Parekh RB, Rohlf C;

XX XX WPI; 2001-557937/62.

XX PT Screening, diagnosis or prognosis of vascular dementia (VD), useful for determining stage of VD and monitoring the effect of VD therapy, for comprises analyzing body fluid by 2-dimensional electrophoresis for features correlated with VD.

XX PS Claim 6; Page 32; 151pp; English.

XX CC The invention relates to screening, diagnosis or prognosis of Vascular Dementia (VD) in a subject comprising analysing body fluid from the subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of features containing at least one chosen feature whose relative abundance correlates with the presence, absence, stage or severity of VD or

CC predicts the onset or course of VD, especially detecting in a sample of cerebrospinal fluid (CSF) from the subject one of 223 VD-associated protein isoforms (VPIs) (ABB5801-ABB56295) as fully defined in the specification. Detecting VD-associated features and VPI is useful for the screening, diagnosis or prognosis of VD, for determining the stage or severity of VD, for identifying a subject at risk of VD or for monitoring the effect of therapy administered to a subject having VD. Nucleic acids encoding a VPI or inhibiting the function of a VPI are useful for the treatment of VD and for gene therapy

XX SQ Sequence 14 AA;

Query Match 2.0%; Score 14; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 6.5e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 619 GSIQVDCBELVSGR 632
|||||
1 GSIQVDCBELVSGR 14

RESULT 8

ABB5910

ID ABB55910 standard; peptide; 10 AA.

XX AC ABB55910;

XX DT 15-FEB-2002 (first entry)

XX DE Vascular dementia-associated protein isoform (VPI) 110.

XX KW Vascular Dementia; VD; VD-associated protein isoform; VPI; screening; diagnosis; prognosis; gene therapy.

XX OS Homo sapiens.

XX PN W0200169261-A2.

XX PD 20-SEP-2001.

XX PF 14-MAR-2001; 2001WO-GB001106.

XX PR 15-MAR-2000; 2000GB-00006285.

XX PR 24-NOV-2000; 2000GB-00028734.

XX PR 28-NOV-2000; 2000US-00724391.

XX PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.

XX PI Herath HWAC, Parekh RB, Rohlf C;

XX XX WPI; 2001-557937/62.

XX PT Screening, diagnosis or prognosis of vascular dementia (VD), useful for determining stage of VD and monitoring the effect of VD therapy, for comprises analyzing body fluid by 2-dimensional electrophoresis for features correlated with VD.

XX PS Claim 6; Page 32; 151pp; English.

XX CC The invention relates to screening, diagnosis or prognosis of Vascular Dementia (VD) in a subject comprising analysing body fluid from the subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of features containing at least one chosen feature whose relative abundance correlates with the presence, absence, stage or severity of VD or predicts the onset or course of VD, especially detecting in a sample of cerebrospinal fluid (CSF) from the subject one of 223 VD-associated protein isoforms (VPIs) (ABB5801-ABB56295) as fully defined in the specification. Detecting VD-associated features and VPI is useful for the screening, diagnosis or prognosis of VD, for determining the stage or severity of VD, for identifying a subject at risk of VD or for monitoring the effect of therapy administered to a subject having VD. Nucleic acids encoding a VPI or inhibiting the function of a VPI are useful for the treatment of VD and for gene therapy

```

XX SQ Sequence 10 AA;
Query Match 1.4%; Score 10; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.041;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 662 FSSGITGCVK 671
Db 1 FSSGITGCVK 10

RESULT 9
ABB55909
ID ABB55909 standard; peptide; 15 AA.
XX AC ABB55909;
XX DT 15-FEB-2002 (first entry)
XX DE Vasculer dementia-associated protein isoform (VPI) 109.
XX KW Vasculer Dementia; VD; VD-associated protein isoform; VPI; screening;
XX KW diagnosis; prognosis; gene therapy.
XX OS Homo sapiens.
XX FN WO200169261-A2.
XX PD 20-SEP-2001.
XX PF 14-MAR-2001; 2001WO-GB001106.
XX PR 15-MAR-2000; 2000GB-00006285.
XX PR 24-NOV-2000; 2000GB-00028734.
XX PR 28-NOV-2000; 2000US-00724391.
XX PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX PI Herath HMAC, Parekh RB, Rohlf C;
XX DR WPI; 2001-557937/62.
XX FT Screening, diagnosis or prognosis of vascular dementia (VD), useful for
XX FT determining stage of VD and monitoring the effect of VD therapy, for
XX FT comprises analyzing body fluid by 2-dimensional electrophoresis for
XX FT features correlated with VD.
XX PS Claim 6; Page 32; 151pp; English.
XX CC The invention relates to screening, diagnosis or prognosis of Vascular
XX CC Dementia (VD) in a subject comprising analysing body fluid from the
XX CC subject by 2-dimensional (2-D) electrophoresis to generate a 2-D array of
XX CC features containing at least one chosen feature whose relative abundance
XX CC correlates with the presence, absence, stage or severity of VD or
XX CC predicts the onset or course of VD, especially detecting in a sample of
XX CC cerebrospinal fluid (CSF) from the subject one of 223 VD-associated
XX CC protein isoforms (VPIs) (ABB55801-ABB56295) as fully defined in the
XX CC specification. Detecting VD-associated features and VPI is useful for the
XX CC screening, diagnosis or prognosis of VD, for determining the stage or
XX CC severity of VD, for identifying a subject at risk of VD or for monitoring
XX CC the effect of therapy administered to a subject having VD. Nucleic acids
XX CC encoding a VPI or inhibiting the function of a VPI are useful for the
XX CC treatment of VD and for gene therapy
XX SQ Sequence 15 AA;
Query Match 1.3%; Score 9; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 587 YQLSGEAR 595
Db 1 YQLSGEAR 9
XX AC ABB65732;
XX DT 20-OCT-2003 (first entry)
XX DE Propionibacterium acnes immunogenic polypeptide #30408.
XX KW Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
XX KW immunostimulant; immune response; vaccine; immunogenic.
XX OS Propionibacterium acnes.
XX FN WO2003033515-A1.
XX PD 24-APR-2003.
XX PF 11-OCT-2002; 2002WO-US032727.
XX PR 15-OCT-2001; 2001US-00978825.
XX PA (CORI-) CORIXA CORP.
XX PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
XX PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
XX PI Barth B, Vallieve-Douglas J;
XX DR WPI; 2003-381789/36.
XX FT New Propionibacterium acnes polypeptides and polynucleotides encoding the
XX FT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
XX FT or for stimulating an immune response specific for a P. acnes protein.
XX PS Claim 7; SEQ ID NO 30408; 1481pp; English.
XX CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
XX CC encoding a Propionibacterium acnes protein. The invention also relates to
XX CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
XX CC immunogenic fragments of P. acnes polypeptides. The invention
XX CC additionally encompasses expression vectors and host cells comprising a
XX CC polynucleotide of the invention; antibodies against polypeptides of the
XX CC invention; fusion proteins comprising a polypeptide of the invention; a
XX CC method for stimulating an immune response specific for a P. acnes
XX CC polypeptide and an isolated T cell population comprising T cells prepared
XX CC via this method; a vaccine composition (comprising P. acnes polypeptides,
XX CC polynucleotides, antibodies, fusion proteins, T cell populations, or
XX CC antigen-presenting cells that express the polypeptide); a method and kit
XX CC for detecting or determining the presence or absence of P. acnes in a
XX CC patient; and a method for inhibiting the development of P. acnes in a
XX CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
XX CC proteins, T cell populations or antigen-presenting cells that express the
XX CC polypeptides are useful for diagnosing, preventing or treating acne
XX CC vulgaris, or for stimulating an immune response specific for a P. acnes
XX CC protein. The polynucleotides can also be used as probes or primers for
XX CC nucleic acid hybridisation. The vaccine composition is useful for the
XX CC stimulation of an immune response against P. acnes, or for treating acne,
XX CC and the kit is useful for performing a diagnostic assay. The present
XX CC sequence represents a specifically claimed P. acnes polypeptide which is
XX CC thought to contain an immunogenic region. Note: The sequence data for
XX CC this patent did not form part of the printed specification, but was
XX CC obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 159 AA;
Query Match 1.3%; Score 9; DB 6; Length 159;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Query Match 1.3%; Score 9; DB 7; Length 2531;
 Best Local Similarity 100.0%; Fred. No. 74;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGTCQ 474
 Db 1029 PCLHGTCQ 1037

RESULT 12
 ADE63705
 ID ADE63705 standard; protein; 2531 AA.

XX AD663705;
 DT 29-JAN-2004 (first entry)
 DE Rat Protein CAA40667, SEQ ID NO 9649.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

OS Rattus norvegicus.
 PN WO2003016475-A2.

XX 27-FEB-2003.
 XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.

XX (GEHO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-269312/26.
 DR GENBANK; CAA40667.

XX New composition comprising two or more isolated polypeptides, useful for
 preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

The invention discloses a composition comprising two or more isolated rat
 or human polynucleotides or a polynucleotide which represents a fragment,
 derivative or allelic variation of the nucleic acid sequence. Also
 claimed are a vector comprising the novel polynucleotide, a host cell
 comprising the vector, a method for identifying a nucleotide sequence
 which is differentially regulated in an animal subjected to pain and a
 kit to perform the method, an array, a method for identifying an agent
 that increases or decreases the expression of the polynucleotide sequence
 that is differentially expressed in neuronal tissue of a first animal
 subjected to pain, a method for identifying a compound which regulates
 the expression of a polynucleotide sequence which is differentially
 expressed in an animal subjected to pain, a method for identifying a
 compound that regulates the activity of one or more of the
 polynucleotides, a method for producing a pharmaceutical composition, a
 method for identifying a compound or small molecule that regulates the
 activity in an animal of one or more of the polypeptides given in the
 specification, a method for identifying a compound useful in treating
 pain and a pharmaceutical composition comprising the one or more
 polypeptides or their antibodies. The polynucleotide or the compound that
 modulates its activity is useful for preparing a medicament for treating
 pain (e.g. spinal segmental nerve injury (SNI), chronic constriction
 injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 therapy). The sequence presented is a rat protein (shown in Table 2 of
 the specification) which is differentially expressed during pain. Note:
 the sequence data for this patent did not form part of the printed

QY 548 VRTSTASGL 556
 Db 31 VRTSTASGL 39

RESULT 11
 ADE63713
 ID ADE63713 standard; protein; 2531 AA.

XX AC ADE63713;
 DT 29-JAN-2004 (first entry)
 DE Rat Protein CAA40667, SEQ ID NO 9657.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

OS Rattus norvegicus.
 PN WO2003016475-A2.

XX 27-FEB-2003.
 XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.

XX (GEHO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-269312/26.
 DR GENBANK; CAA40667.

XX New composition comprising two or more isolated polypeptides, useful for
 preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

The invention discloses a composition comprising two or more isolated rat
 or human polynucleotides or a polynucleotide which represents a fragment,
 derivative or allelic variation of the nucleic acid sequence. Also
 claimed are a vector comprising the novel polynucleotide, a host cell
 comprising the vector, a method for identifying a nucleotide sequence
 which is differentially regulated in an animal subjected to pain and a
 kit to perform the method, an array, a method for identifying an agent
 that increases or decreases the expression of the polynucleotide sequence
 that is differentially expressed in neuronal tissue of a first animal
 subjected to pain, a method for identifying a compound which regulates
 the expression of a polynucleotide sequence which is differentially
 expressed in an animal subjected to pain, a method for identifying a
 compound that regulates the activity of one or more of the
 polynucleotides, a method for producing a pharmaceutical composition, a
 method for identifying a compound or small molecule that regulates the
 activity in an animal of one or more of the polypeptides given in the
 specification, a method for identifying a compound useful in treating
 pain and a pharmaceutical composition comprising the one or more
 polypeptides or their antibodies. The polynucleotide or the compound that
 modulates its activity is useful for preparing a medicament for treating
 pain (e.g. spinal segmental nerve injury (SNI), chronic constriction
 injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 therapy). The sequence presented is a rat protein (shown in Table 2 of
 the specification) which is differentially expressed during pain. Note:
 the sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic form directly from WIPO at
 ftp.wipo.int/pub/published_pt_sequences.

CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2531 AA;

Query Match 1.3%; Score 9; DB 7; Length 2531;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 466 PCLHGTCQ 474
 Db 1029 PCLHGTCQ 1037

RESULT 13
 ADE63709
 ID ADE63709 standard; protein; 2531 AA.

AC ADE63709;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Rat Protein CAA40667, SEQ ID NO 9653.
 DE
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 KW
 XX Rattus norvegicus.
 XX
 OS WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 PF 14-AUG-2002; 2002WO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 XX (GENO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 PA
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX WPI; 2003-268312/26.
 DR GENBANK; CAA40667.
 XX

PT New composition comprising two or more isolated polypeptides, useful for
 DR preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition,
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction

CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2531 AA;

Query Match 1.3%; Score 9; DB 7; Length 2531;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 466 PCLHGTCQ 474
 Db 1029 PCLHGTCQ 1037

RESULT 14
 ADE63701
 ID ADE63701 standard; protein; 2531 AA.

AC ADE63701;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Rat Protein CAA40667, SEQ ID NO: 9645.
 DE
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 KW
 XX Rattus norvegicus.
 XX
 OS WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 PF 14-AUG-2002; 2002WO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 XX (GENO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 PA
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX WPI; 2003-268312/26.
 DR GENBANK; CAA40667.
 XX

PT New composition comprising two or more isolated polypeptides, useful for
 DR preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition,
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating

CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2531 AA;

Query Match 1.3%; Score 9; DB 7; Length 2531;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 466 PCLHGTCQ 474
 |||||
 Db 1029 PCLHGTCQ 1037

RESULT 15
 ABB70376
 ID ABB70376 standard; protein; 3319 AA.

AC ABB70376;
 XX
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 37920.
 XX
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW
 XX Drosophila melanogaster.
 OS
 XX
 XX WO200171042-A2.
 PN
 XX
 PD 27-SEP-2001.
 XX
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 XX 23-MAR-2000; 2000US-0191637P.
 PR
 PR 11-JUL-2000; 2000US-00614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX
 FI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656960/75.
 DR
 DR N-PSDB; ABL14479.
 XX
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 XX Disclosure; SEQ ID NO 37920; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABLJ30511), expressed DNA
 CC sequences (ABU01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 3319 AA;

Query Match 1.3%; Score 9; DB 4; Length 3319;

Best Local Similarity 100.0%; Pred. No. 96;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 603 NDGEHHRVT 611
 |||||
 Db 3033 NDGEHHRVT 3041

Search completed: March 9, 2004, 17:27:53
 Job time : 88.8465 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: March 9, 2004, 17:25:24 ; Search time 27.0475 Seconds
(without alignments)
1345.642 Million cell updates/sec

Title: US-10-006-011A-3
Perfect score: 705
Sequence: 1 EKKITFRPDSADGMLLYNQ.....QLDLQHRQAQANTRCPS 705

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfilesi.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Table with columns: Result No., Score, Query Match, Length, DB ID, Description. Contains 27 rows of search results.

Table with columns: 28-45, 7-7, 1.0-1.0, 100-290, 4-4, US-09-363-316B-3 to US-09-252-991A-27099

ALIGNMENTS

RESULT 1
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; TITLE OF INVENTION: AND LIGANDS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESS: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/644,271
; FILING DATE: 10-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 60/008,657
; FILING DATE: 15-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cobert Robert J
; REGISTRATION NUMBER: 36,108
; REFERENCE/DOCKET NUMBER: REG 195A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
; TELEX:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1940 amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Rat Agrin
; LOCATION: 1...1940
; OTHER INFORMATION:
; US-08-644-271-30
Query Match 1.3%; Score 9; DB 2; Length 1940;

Best Local Similarity 100.0%; Pred. No. 20; Mismatches 0; Indels 0; Gaps 0;

466 PCLHGTCQ 474
1228 PCLHGTCQ 1236

RESULT 2
US-09-077-955-34
Sequence 34, Application US/09077955A
Patent No. 6413740
GENERAL INFORMATION:
APPLICANT: Valenzuela et al., David M.
TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
FILE REFERENCE: REG195-B-PCT-US
CURRENT APPLICATION NUMBER: US/09/077,955A
EARLIER FILING DATE: 1998-09-10
EARLIER APPLICATION NUMBER: PCT/US96/20696
EARLIER FILING DATE: 1996-12-13
EARLIER APPLICATION NUMBER: 08/644,271
EARLIER FILING DATE: 1996-05-10
EARLIER APPLICATION NUMBER: 60/008,657
EARLIER FILING DATE: 1995-12-15
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 34
LENGTH: 1940
TYPE: PRT
ORGANISM: Rattus sp.
US-09-077-955-34

Query Match 1.3%; Score 9; DB 4; Length 1940;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

466 PCLHGTCQ 474
1228 PCLHGTCQ 1236

RESULT 3
US-107-532A-5434
Sequence 5434, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER TYPE: CD/ROM ISO9660
MEDIUM TYPE: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5434:
SEQUENCE CHARACTERISTICS:
LENGTH: 241 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:

NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...241
SEQUENCE DESCRIPTION: SEQ ID NO: 5434:
US-09-107-532A-5434

Query Match 1.1%; Score 8; DB 4; Length 241;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

144 IVFHDLNL 151
28 IVFHDLNL 35

RESULT 4
US-09-252-991A-19300
Sequence 19300, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 19300
LENGTH: 427
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19300

Query Match 1.1%; Score 8; DB 4; Length 427;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

76 LLRSLTQG 83
247 LLRSLTQG 254

RESULT 5
US-09-540-236-2071
Sequence 2071, Application US/09540236
Patent No. 6873510
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2005-001
CURRENT APPLICATION NUMBER: US/09/540,236
CURRENT FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 2071
LENGTH: 579

```

; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2071

```

```

Query Match 1.1%; Score 8; DB 4; Length 579;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 319 AEPALGR 326
Db 320 AEPALGR 327

```

```

RESULT 6
US-08-537-210A-1
; Sequence 1, Application US/08537210A
; Patent No. 5780300
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Fortini, Mark
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,210A
FILING DATE: 29-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Miscock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-027
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1015 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: hum N (Human No. 5780300ch 2)
LOCATION: 1155...2169
OTHER INFORMATION: Highly conserved ankyrin repeat
US-08-537-210A-1

```

```

Query Match 1.1%; Score 8; DB 1; Length 1015;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 426 PCQHGATC 433
Db 3 PCQHGATC 10

```

```

RESULT 7
US-09-113-825-1

```

```

; Sequence 1, Application US/09113825
; Patent No. 6149902
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Fortini, Mark
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/113,825
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/537,210
FILING DATE: 29-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: Miscock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-027
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1015 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: hum N (Human No. 6149902ch 2)
LOCATION: 1155...2169
OTHER INFORMATION: Highly conserved ankyrin repeat
US-09-113-825-1

```

```

Query Match 1.1%; Score 8; DB 3; Length 1015;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 426 PCQHGATC 433
Db 3 PCQHGATC 10

```

```

RESULT 8
US-08-460-309-2
; Sequence 2, Application US/08460309
; Patent No. 5837496
GENERAL INFORMATION:
APPLICANT: Engvall, Eva
APPLICANT: Leivo, Ilmo
TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
TITLE OF INVENTION: Fragments and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego

```

```

; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/460,309
; FILING DATE:
; PRIORITY APPLICATION DATA:
; CLASSIFICATION: 435
; PRIORITY APPLICATION NUMBER: US 08/125,077
; FILING DATE: 22-SEP-1993
; APPLICATION NUMBER: US PCT/US 94/10730
; FILING DATE: 21-SEP-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/472,319
; FILING DATE: 27-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 9721
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-8949
; TELEFAX: (619) 535-9001
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1130 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-460-309-2

```

```

Query Match 1.1%; Score 8; DB 2; Length 1130;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 543 TIELEVRT 550
Db 808 TIELEVRT 815

```

```

RESULT 10
5444158-2
; Patent No. 5444158
; APPLICANT: ENGVALL, EVA; SANES, JOSHUA
; TITLE OF INVENTION: MEROSIN, NUCLEIC ACIDS ENCODING,
; FRAGMENTS AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/97,642
; FILING DATE: 08-JUL-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 587,689
; FILING DATE: 24-SEP-1990
; APPLICATION NUMBER: 472,319
; FILING DATE: 30-JAN-1990
; SEQ ID NO: 2:
; LENGTH: 1130
5444158-2

```

```

Query Match 1.1%; Score 8; DB 6; Length 1130;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 543 TIELEVRT 550
Db 808 TIELEVRT 815

```

```

RESULT 11
US-08-340-428B-49
; Sequence 49 Application US/08340428B
; Patent No. 5648465
; GENERAL INFORMATION:
; APPLICANT: MARGOLIS, Richard U.
; APPLICANT: RAUCH, Uwe
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
; TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
; NUMBER OF SEQUENCES: 49

```

```

; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/125,077
; FILING DATE: 22-SEP-1993
; ADDRESS: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/125,077
; FILING DATE: 22-SEP-1993

```

/ CORRESPONDENCE ADDRESS:
 / ADDRESS: Browdy and Neimark
 / STREET: 419 Seventh Street, N.W.
 / CITY: Washington
 / STATE: D.C.
 / COUNTRY: U.S.A.
 / ZIP: 20004
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patent In Release #1.0, Version #1.25
 / CURRENT APPLICATION DATA:
 / FILING DATE: 14 No. 5648465ember 1994
 / CLASSIFICATION: 514
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: 07/922,911
 / FILING DATE: 03 August 1992
 / CLASSIFICATION: 514
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Browdy, Roger L.
 / REGISTRATION NUMBER: 25,618
 / REFERENCE/DOCKET NUMBER: Margolis-1A
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: 202-628-5197
 / TELEFAX: 202-737-3528
 / INFORMATION FOR SEQ ID NO: 49:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 1257 amino acids
 / TYPE: amino acid
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / MOLECULE TYPE: peptide
 / US-08-340-428B-49

Query Match 1.1%; Score 8; DB 1; Length 1257;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGTC 473
 Db 957 PCLRGTC 964

RESULT 12
 US-09-230-652-2
 ; Sequence 2, Application US/09230652A
 ; GENERAL INFORMATION:
 ; APPLICANT: Tournier-Lasserve, Elisabeth
 ; APPLICANT: Joutel, Anne
 ; APPLICANT: Bousser, Marie-Germaine
 ; APPLICANT: Bach, Jean-Francois
 ; TITLE OF INVENTION: GENE INVOLVED IN CADASIL, METHOD OF DIAGNOSIS AND
 ; FILE REFERENCE: 03715.0048-00000
 ; CURRENT APPLICATION NUMBER: US/09/230.652A
 ; CURRENT FILING DATE: 1999-05-17
 ; EARLIER APPLICATION NUMBER: FR 96 09733
 ; EARLIER FILING DATE: 1996-08-01
 ; EARLIER APPLICATION NUMBER: FR 97 04680
 ; EARLIER FILING DATE: 1997-04-16
 ; EARLIER APPLICATION NUMBER: PCT/FR97/01433
 ; EARLIER FILING DATE: 1997-07-31
 ; NUMBER OF SEQ ID NOS: 163
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 2321
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: human ADNC No. 6537775ch 3

US-09-230-652-2
 Query Match 1.1%; Score 8; DB 4; Length 2321;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 GFGPRCQ 490
 Db 1237 GFGPRCQ 1244

RESULT 13
 US-08-185-432-16
 ; Sequence 16, Application US/08185432
 ; Patent No. 5750652
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, Spyridon
 ; APPLICANT: Busseau, Isabelle
 ; APPLICANT: Diederich, Robert J.
 ; APPLICANT: Xu, Tian
 ; APPLICANT: Matsuno, Kenji
 ; TITLE OF INVENTION: DELTEx PROTEINS, NUCLEIC ACIDS, AND
 ; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: PENNIE & EDWARDS
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/185,432
 ; FILING DATE: 21-JAN-1994
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mistock, S. Leslie
 ; REGISTRATION NUMBER: 18,872
 ; REFERENCE/DOCKET NUMBER: 7326-006
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-8864/9741
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2471 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 / US-08-185-432-16

Query Match 1.1%; Score 8; DB 1; Length 2471;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 426 PCQHGATC 433
 Db 1157 PCQHGATC 1164

RESULT 14
 US-08-083-590A-19
 ; Sequence 19, Application US/08083590A
 ; Patent No. 5786158
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, S. et al.
 ; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
 ; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And

```

; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/3741
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2471 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-083-590A-19

```

```

Query Match 1.1%; Score 8; DB 1; Length 2471;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 426 PCQHGATC 433
Db 1157 PCQHGATC 1164

```

```

RESULT 15
US-08-532-384-19
; Sequence 19, Application US/08532384
; Patent No. 6083904
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,384
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,590
; FILING DATE: 23-JUN-1993

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/3741
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2471 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-532-384-19

```

```

Query Match 1.1%; Score 8; DB 3; Length 2471;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 426 PCQHGATC 433
Db 1157 PCQHGATC 1164

```

```

Search completed: March 9, 2004, 17:32:24
Job time: 28.0475 secs

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:27:59 ; Search time 51.7431 Seconds
(without alignments)
2876.963 Million cell updates/sec

Title: US-10-006-011A-3

Perfect score: 705

Sequence: 1 EKIFRPSADGMLLYNQ.....QLDLQHRQAQANTRPCPS 705

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 809742 seqs, 211153259 residues

Word size : 0

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : Published Applications_AA:*
- 1: /cgn2_6/ptodata1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata1/pubpaa/US09A_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata1/pubpaa/US09B_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata1/pubpaa/US09C_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata1/pubpaa/US09_NEW_PUB.pep.*
 - 12: /cgn2_6/ptodata1/pubpaa/US10A_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata1/pubpaa/US10B_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata1/pubpaa/US10C_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata1/pubpaa/US10_NEW_PUB.pep.*
 - 16: /cgn2_6/ptodata1/pubpaa/US60_NEW_PUB.pep.*
 - 17: /cgn2_6/ptodata1/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1.3	1940	13	US-10-016-283-34	Sequence 34, Appl
2	1.3	2531	15	US-10-190-115-29	Sequence 29, Appl
3	1.3	2531	15	US-10-369-072-29	Sequence 29, Appl
4	1.1	54	10	US-09-776-724A-136	Sequence 136, Appl
5	1.1	80	9	US-09-864-761-35784	Sequence 35784, A
6	1.1	131	15	US-10-264-237-2547	Sequence 2547, Ap
7	1.1	264	11	US-09-833-245-2022	Sequence 2022, Ap
8	1.1	264	15	US-10-266-829-75	Sequence 75, Appl
9	1.1	299	14	US-10-219-220-251	Sequence 251, Appl
10	1.1	309	14	US-10-122-706-31	Sequence 31, Appl
11	1.1	383	11	US-09-833-245-770	Sequence 770, Appl
12	1.1	383	14	US-10-245-103-80	Sequence 80, Appl
13	1.1	383	14	US-10-245-107-80	Sequence 80, Appl
14	1.1	383	14	US-10-245-143-80	Sequence 80, Appl
15	1.1	383	14	US-10-245-771-80	Sequence 80, Appl

16	1.1	383	14	US-10-245-851-80	Sequence 80, Appl
17	1.1	383	14	US-10-245-883-80	Sequence 80, Appl
18	1.1	383	14	US-10-237-535-80	Sequence 80, Appl
19	1.1	383	14	US-10-238-183-80	Sequence 80, Appl
20	1.1	383	14	US-10-238-283-80	Sequence 80, Appl
21	1.1	383	14	US-10-238-370-80	Sequence 80, Appl
22	1.1	383	14	US-10-245-055-80	Sequence 80, Appl
23	1.1	383	14	US-10-245-147-80	Sequence 80, Appl
24	1.1	383	14	US-10-245-730-80	Sequence 80, Appl
25	1.1	383	14	US-10-245-739-80	Sequence 80, Appl
26	1.1	383	14	US-10-246-210-80	Sequence 80, Appl
27	1.1	383	14	US-10-239-196-80	Sequence 80, Appl
28	1.1	383	14	US-10-243-024-80	Sequence 80, Appl
29	1.1	383	14	US-10-243-409-80	Sequence 80, Appl
30	1.1	383	14	US-10-245-621-80	Sequence 80, Appl
31	1.1	383	14	US-10-245-880-80	Sequence 80, Appl
32	1.1	383	14	US-10-245-033-80	Sequence 80, Appl
33	1.1	383	14	US-10-243-095-80	Sequence 80, Appl
34	1.1	383	14	US-10-245-185-80	Sequence 80, Appl
35	1.1	383	14	US-10-245-427-80	Sequence 80, Appl
36	1.1	383	14	US-10-245-473-80	Sequence 80, Appl
37	1.1	383	14	US-10-245-770-80	Sequence 80, Appl
38	1.1	383	14	US-10-245-877-80	Sequence 80, Appl
39	1.1	383	14	US-10-246-976-80	Sequence 80, Appl
40	1.1	383	14	US-10-243-320-80	Sequence 80, Appl
41	1.1	383	14	US-10-242-743-80	Sequence 80, Appl
42	1.1	383	14	US-10-242-845-80	Sequence 80, Appl
43	1.1	383	14	US-10-237-636-80	Sequence 80, Appl
44	1.1	383	14	US-10-238-325-80	Sequence 80, Appl
45	1.1	383	14	US-10-238-346-80	Sequence 80, Appl

ALIGNMENTS

RESULT 1
US-10-016-283-34
; Sequence 34, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/10/016,283
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US/09/077,955A
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: PCT/US96/20696
; PRIOR FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 34
; LENGTH: 1940
; TYPE: PRT
; ORGANISM: Rattus sp.
; US-10-016-283-34

Query Match 1.3%; Score 9; DB 13; Length 1940;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 465 PCLHGGTQC 474
Db 1228 PCLHGGTQC 1236
RESULT 2
US-10-190-115-29
; Sequence 29, Application US/10190115
; Publication No. US20030207394A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrock, John P. II
; APPLICANT: Boldog, Ferenc L.

```

; APPLICANT: Burgess, Catherine E.
; APPLICANT: Caeman, Stacie J.
; APPLICANT: Grosse, William M.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Ji, Weizhen
; APPLICANT: Lepley, Denise M.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Mezick, Amanda J.
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: Shen, Lei
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Szerkes, Edward S. Jr.
; APPLICANT: Taupier, Raymond J. Jr.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Voss, Edward Z.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-050 CIP
; CURRENT APPLICATION NUMBER: US/10/190,115
; CURRENT FILING DATE: 2003-02-10
; PRIOR APPLICATION NUMBER: 60/303,168
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/368,996
; PRIOR FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 60/386,816
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: 60/215,854
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,856
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/216,585
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,722
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,622
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,586
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,622
; PRIOR FILING DATE: 2000-07-07
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 136
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 29
; LENGTH: 2531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-190-115-29

```

```

Query Match 1.3%; Score 9; DB 15; Length 2531;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 466 PCLHGTCQ 474
Db 1029 PCLHGTCQ 1037

RESULT 3
US-10-369-072-29
; Sequence 29, Application US/10369072
; Publication No. US20040014081A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Spaderna, Stephen K
; APPLICANT: Tchernev, Velizar
; APPLICANT: Liu, Xiaohong
; APPLICANT: Shenoy, Suresh

```

```

; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Taupier, Raymond T
; APPLICANT: Rastelli, Luca
; APPLICANT: Grosse, William M
; APPLICANT: Szerkes, Edward S
; APPLICANT: Lepley, Denise M
; APPLICANT: Shen, Lei
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Padigar, Muralidhara
; TITLE OF INVENTION: No. US20040014081A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-050 CONZ
; CURRENT APPLICATION NUMBER: US/10/369,072
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 10/174,372
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 09/898,994
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/215,854
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,856
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,902
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/216,585
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,586
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,722
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/218,622
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/218,992
; PRIOR FILING DATE: 2000-07-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 2531
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-369-072-29

```

```

Query Match 1.3%; Score 9; DB 15; Length 2531;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 466 PCLHGTCQ 474
Db 1029 PCLHGTCQ 1037

RESULT 4
US-09-776-724A-136
; Sequence 136, Application US/09776724A
; Publication No. US20030050455A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 64 Human Secreted Proteins
; FILE REFERENCE: P2011
; CURRENT APPLICATION NUMBER: US/09/776,724A
; CURRENT FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: 60/180,909
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 09/669,688
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: 09/229,982
; PRIOR FILING DATE: 1999-01-14
; PRIOR APPLICATION NUMBER: PCT/US98/14613
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/052,661

```

```

; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/052,872
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/052,871
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/052,874
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/052,873
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/052,870
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/052,875
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: 60/053,440
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,441
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,442
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/056,359
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,725
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,985
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,952
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,989
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/056,361
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,726
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,724
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,946
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/055,683
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 136
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (32)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (39)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-776-724A-136

```

```

Query Match 1.1%; Score 8; DB 10; Length 54;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 74 VTLRLSLT 81
Db 47 VTLRLSLT 54

```

```

RESULT 5
US-09-864-761-35784
; Sequence 35784, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Weisheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

```

```

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecom,ca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35784
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035413.19
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
; OTHER INFORMATION: EST_HUMAN HT: BE875511.1, EVALUATE 3.00e-14
; US-09-864-761-35784

```

```

Query Match 1.1%; Score 8; DB 9; Length 80;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 680 FGAPPPQP 687
Db 33 FGAPPPQP 40

```

```

RESULT 6
US-10-264-237-2547
; Sequence 2547, Application US/10264237

```


Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P4131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2547
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (108)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (106)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (115)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (120)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (127)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-2547

Query Match 1.1%; Score 8; DB 15; Length 131;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 GYLLFSGG 285
Db 58 GYLLFSGG 65

RESULT 7

US-09-833-245-2022
; Sequence 2022, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2022
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-2022

Query Match 1.1%; Score 8; DB 11; Length 264;

Best Local Similarity 100.0%; Pred. No. 66;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 GYLLFSGG 285
Db 13 GYLLFSGG 20

RESULT 8

US-10-266-829-75
; Sequence 75, Application US/10266829
; Publication No. US20030220489A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 29 Human secreted proteins
; FILE REFERENCE: PZ041P1
; CURRENT APPLICATION NUMBER: US/10/266,829
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: 09/756,168
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: PCT/US00/19735
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 60/145,220
; PRIOR FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 75
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-266-829-75

Query Match 1.1%; Score 8; DB 15; Length 264;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 GYLLFSGG 285
Db 13 GYLLFSGG 20

RESULT 9

US-10-219-220-251
; Sequence 251, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE REFERENCE: 11000.1022c1
; CURRENT APPLICATION NUMBER: US/10/219,220
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 251
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-219-220-251

Query Match 1.1%; Score 8; DB 14; Length 299;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 VTLRLSLT 81
Db 260 VTLRLSLT 267

RESULT 10

US-10-122-706-31
 ; Sequence 31, Application US/10122706
 ; Publication No. US20030119012A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Srinivasan, Maithreyan
 ; APPLICANT: Reifler, Michael
 ; TITLE OF INVENTION: Sulfurylase-Luciferase Fusion Proteins
 ; FILE REFERENCE: 21465-504
 ; CURRENT APPLICATION NUMBER: US/10/122,706
 ; CURRENT FILING DATE: 2002-07-01
 ; PRIOR APPLICATION NUMBER: 60/335,949
 ; PRIOR FILING DATE: 2001-10-30
 ; NUMBER OF SEQ ID NOS: 31
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 31
 ; LENGTH: 309
 ; TYPE: PRT
 ; ORGANISM: Thermomonospora fusca
 US-10-122-706-31

Query Match 1.1%; Score 8; DB 14; Length 309;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 VLLFSGGK 286
 |||||
 Db 38 VLLFSGGK 45

RESULT 11
 US-09-833-245-770
 ; Sequence 770, Application US/09833245
 ; Publication No. US20040010134A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences, Inc.
 ; TITLE OF INVENTION: Albumin Fusion Proteins
 ; FILE REFERENCE: PF546PCT
 ; CURRENT APPLICATION NUMBER: US/09/833,245
 ; CURRENT FILING DATE: 2001-04-12
 ; PRIOR APPLICATION NUMBER: 60/229, 358
 ; PRIOR FILING DATE: 2000-04-12
 ; PRIOR APPLICATION NUMBER: 60/256, 931
 ; PRIOR FILING DATE: 2000-12-21
 ; PRIOR APPLICATION NUMBER: 60/199, 384
 ; PRIOR FILING DATE: 2000-04-25
 ; NUMBER OF SEQ ID NOS: 2267
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 770
 ; LENGTH: 383
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-833-245-770

Query Match 1.1%; Score 8; DB 11; Length 383;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 PCQGGQC 173
 |||||
 Db 100 PCQGGQC 107

RESULT 12
 US-10-245-103-80
 ; Sequence 80, Application US/10245103
 ; Publication No. US20030068779A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin
 ; APPLICANT: Eaton, Dan
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Grimaldi, J. Christopher
 ; APPLICANT: Gurney, Austin

APPLICANT: Smith, Victoria
 APPLICANT: Stephan, Jean-Phillippe
 APPLICANT: Watanabe, Colin
 APPLICANT: Wood, William
 APPLICANT: Zhang, Zemin
 APPLICANT: Fong, Sherman
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 TITLE OF INVENTION: ACIDS ENCODING THE SAME
 FILE REFERENCE: P3630R1C112
 CURRENT APPLICATION NUMBER: US/10/245,103
 CURRENT FILING DATE: 2002-09-17
 PRIOR APPLICATION NUMBER: 10/197942
 PRIOR FILING DATE: 2002-07-18
 PRIOR APPLICATION NUMBER: 60/059114
 PRIOR FILING DATE: 1997-09-17
 PRIOR APPLICATION NUMBER: 60/063046
 PRIOR FILING DATE: 1997-10-24
 PRIOR APPLICATION NUMBER: 60/065027
 PRIOR FILING DATE: 1997-11-10
 PRIOR APPLICATION NUMBER: 60/079689
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/086478
 PRIOR FILING DATE: 1998-05-22
 PRIOR APPLICATION NUMBER: 60/087607
 PRIOR FILING DATE: 1998-06-02
 PRIOR APPLICATION NUMBER: 60/089801
 PRIOR FILING DATE: 1998-06-18
 PRIOR APPLICATION NUMBER: 60/090557
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090689
 PRIOR FILING DATE: 1998-06-25
 Remaining Prior Application data removed - See File Wrapper or PALM.
 NUMBER OF SEQ ID NOS: 116
 SEQ ID NO 80
 LENGTH: 383
 TYPE: PRT
 ORGANISM: Homo Sapien
 US-10-245-103-80

Query Match 1.1%; Score 8; DB 14; Length 383;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 PCQGGQC 173
 |||||
 Db 100 PCQGGQC 107

RESULT 13
 US-10-245-107-80
 ; Sequence 80, Application US/10245107
 ; Publication No. US20030068779A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin
 ; APPLICANT: Eaton, Dan
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Grimaldi, J. Christopher
 ; APPLICANT: Gurney, Austin
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stephan, Jean-Phillippe
 ; APPLICANT: Watanabe, Colin
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; APPLICANT: Fong, Sherman
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 TITLE OF INVENTION: ACIDS ENCODING THE SAME
 FILE REFERENCE: P3630R1C71
 CURRENT APPLICATION NUMBER: US/10/245,107
 CURRENT FILING DATE: 2002-09-16
 PRIOR APPLICATION NUMBER: 10/197942
 PRIOR FILING DATE: 2002-07-18
 PRIOR APPLICATION NUMBER: 60/059114

```

; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 80
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-107-80

```

```

Query Match 1.1%; Score 8; DB 14; Length 383;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 166 PCQNGQC 173
Db 100 PCQNGQC 107

```

```

RESULT 14
US-10-245-143-80
; Sequence 80, Application US/10245143
; Publication No. US2003068780A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C90
; CURRENT APPLICATION NUMBER: US/10/245,143
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 80
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-107-80

```

```

; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 80
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-143-80

```

```

Query Match 1.1%; Score 8; DB 14; Length 383;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 166 PCQNGQC 173
Db 100 PCQNGQC 107

```

```

RESULT 15
US-10-245-771-80
; Sequence 80, Application US/10245771
; Publication No. US2003068781A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C98
; CURRENT APPLICATION NUMBER: US/10/245,771
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 80
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-771-80

```

```

Query Match 1.1%; Score 8; DB 14; Length 383;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 166 PCONGQC 173
| | | | |
Db 100 PCONGQC 107

Search completed: March 9, 2004, 17:34:05
Job time : 52.7431 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:23:39 ; Search time 24.1076 Seconds
 (without alignments)
 2813.016 Million cell updates/sec

Title: US-10-006-011A-3

Perfect score: 705

Sequence: 1 EIKITFRPDSADGMLLYNGQ.....QPLDLQHRAQAGANTRPCPS 705

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 78:**

1: Pir1:**

2: Pir2:**

3: Pir3:**

4: Pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	705	100.0	4391	A38096	perlecan precursor
2	44	6.2	3707	S18252	heparan sulfate pr
3	9	1.3	861	A48825	Notch homolog Motc
4	9	1.3	1959	AGRT	agrin - rat
5	9	1.3	2531	S18188	notch protein homo
6	9	1.3	2531	A46019	notch-1 protein -
7	8	1.1	215	T48789	Tfsl related prote
8	8	1.1	332	A70772	probable sulfate a
9	8	1.1	430	A95335	conserved hypothet
10	8	1.1	497	AH2015	sodium/solute symp
11	8	1.1	500	F70012	leucyl aminopeptid
12	8	1.1	748	T00732	hypothetical prote
13	8	1.1	886	T39081	hypothetical prote
14	8	1.1	1203	A49175	Notch B protein-
15	8	1.1	1257	S28764	neurocan precursor
16	8	1.1	1268	S52781	neurocan - mouse
17	8	1.1	1328	T43060	agrin - electric r
18	8	1.1	1751	MMHUMH	laminin alpha-2 ch
19	8	1.1	1955	AGCH	agrin precursor -
20	8	1.1	2139	A35672	crumbs protein - f
21	8	1.1	2321	S78549	notch3 protein - h
22	8	1.1	2471	A49128	cell-fate determin
23	8	1.1	2531	T31070	notch homolog - se
24	8	1.1	3106	S53868	laminin alpha-2 ch
25	7	1.0	46	S21922	T-cell receptor al
26	7	1.0	93	S75571	hypothetical prote
27	7	1.0	96	E97765	mutator protein Mu
28	7	1.0	129	T29451	hypothetical prote
29	7	1.0	137	T43574	translocation prot

virC-region hypoth
 gene 22 protein -
 hypothetical prote
 hypothetical prote
 hypothetical prote
 hypothetical prote
 phosphotransfere
 hypothetical prote
 probable peptidyl-
 tenebrosin C - sea
 chemotaxis protein
 probable transcrip
 hypothetical prote
 hypothetical prote
 equinatoxin II pre
 hypothetical prote
 photosystem II pro

ALIGNMENTS

RESULT 1
 A38096
 perlecan precursor - human
 N:Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate proteo
 C:Species: Homo sapiens (man)
 C>Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #seq change 05-Nov-1999
 C:Accession: A38096; S19256; S77946; A41059; A40306; B33625; A33625; A41736
 R:Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
 J. Biol. Chem. 267, 8544-8557, 1992
 Article: Primary structure of the human heparan sulfate proteoglycan from basement membr
 tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
 A:Reference number: A38096; MUID:92235084; PMID:1569102
 A:Accession: A38096
 A:Molecule type: mRNA
 A:Residues: 1-4391 <MUR>
 A:Cross-references: GB:M85289; NID:G184426; PIDN:AAA52700.1; PID:G184427
 R:Kallunki, P.; Tryggvason, K.
 J. Cell Biol. 116, 559-571, 1992
 A:Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD prot
 eil adhesion molecules, and epidermal growth factor.
 A:Reference number: A41736; MUID:92112994; PMID:1730768
 A:Accession: S19256
 A:Molecule type: mRNA
 A:Residues: 1-57, D, '59-434, 'A, 436, 'FL', 438-449, 'Q', 451-502, 'A', 503-792, 'K', 794-908, 'R',
 71-2979, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 3633-3:
 A:Cross-references: EMBL:X62515
 R:Tryggvason, K.
 submitted to the EMBL Data Library, October 1991
 A:Reference number: S77946
 A:Accession: S77946
 A:Molecule type: mRNA
 A:Residues: 1-57, 'D', '59-434, 'A, 436, 'FL', 438-449, 'Q', 451-502, 'A', 503-792, 'K', 794-908, 'R',
 71-2979, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 3633-4(
 A:Cross-references: EMBL:X62515; NID:G29469; PIDN:CAA44373.1; PID:G29470
 R:Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
 Genomics 11, 389-396, 1991
 A:Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the g
 A:Reference number: A41059; MUID:92120660; PMID:1685141
 A:Accession: A41059
 A:Molecule type: mRNA
 A:Residues: 'RT', 892-908, 'R', 910-1101, 'L', 1103-1132, 'L', 1134-1221, 'L', 1223-1397 <KA2>
 A:Cross-references: GB:S76436; NID:G243370; PIDN:AAB21121.1; PID:G243371
 R:Dodge, G.R.; Kovalesky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo, F
 Genomics 10, 673-680, 1991
 A:Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellular
 A:Reference number: A40306; MUID:91365376; PMID:1679749
 A:Accession: A40306
 A:Molecule type: mRNA
 A:Residues: 1018-1405, 'G', 1407-1409, 'G', 1411-1465 <DOD>
 A:Cross-references: GB:M64283; NID:G184424; PIDN:AAA52699.1; PID:G184425
 R:Heemans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den
 J. Cell Biol. 109, 3199-3211, 1989

A>Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclonal antibodies.
 A:Reference number: A33625; MUID:90078352; PMID:2687294
 A:Accession: B33625
 A:Molecule type: protein
 A:Residues: 1379-1384,'X',1386-1388,'X',1390-1398 <HE2>
 A:Accession: A33625
 A:Molecule type: protein
 A:Residues: 2166-2171,'X',2173-2175,'X',2177-2185 <HE3>
 A>Note: peptide potentially matches four different regions of sequence shown
 C:Genetics:
 A:Gene: GDB:HSR92
 A:Cross-references: GDB:126372; OMIM:142461
 A:Map position: lp36.1-1p36.1
 C:Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repeat
 C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembrane
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-4391/Product: perlecan #status predicted <MAT>
 F:22-193/Domain: I <DOM1>
 F:194-530/Domain: II <DOM2>
 F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F:531-1676/Domain: III <DOM3>
 F:1159-1206/Domain: laminin-type EGF-like homology <LEG>
 F:1563-1610/Domain: laminin-type EGF-like homology <LEG7>
 F:1613-1668/Domain: laminin-type EGF-like homology <LEG8>
 F:1677-3686/Domain: IV <DOM4>
 F:2007-2034/Domain: transmembrane #status predicted <TRM>
 F:3687-4391/Domain: V <DOM5>
 F:3845-3880/Domain: EGF homology <EGF1>
 F:3888-3921/Domain: EGF homology <EGF>
 F:3993-4106/Domain: laminin G repeat homology <LG2>
 F:4147-4175/Domain: EGF homology <EGF2>
 F:4149-4151/Region: motor neuron attachment (L-R-E) motif
 F:4299-4301/Region: motor neuron attachment (L-R-E) motif
 F:65.71.76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
 F:89.554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (covalent)
 F:2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 705; DB 2; Length 4391;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EIKTFRPSDADGMLLYNGKRVPSPTLNANRQPDFISGLYGRPEFDFAGSGMATI 60
 |||||
 Db 3687 EIKTFRPSDADGMLLYNGKRVPSPTLNANRQPDFISGLYGRPEFDFAGSGMATI 3746

Qy 61 RHPTPLALGHFHTVTLRLSLTQSLIYVGLDLPVNGTSGKRFQGLDLNEELYLGGYDYGA 120
 |||||
 Db 3747 RHPTPLALGHFHTVTLRLSLTQSLIYVGLDLPVNGTSGKRFQGLDLNEELYLGGYDYGA 3806

Qy 121 IPKAGLSGFRIGCVRELRIOGEELVFDLNLTAHGISHCFTCDRPFQNGQCHDSSESS 180
 |||||
 Db 3807 IPKAGLSGFRIGCVRELRIOGEELVFDLNLTAHGISHCFTCDRPFQNGQCHDSSESS 3866

Qy 181 YVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSLRCEGVT 240
 |||||
 Db 3867 YVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSLRCEGVT 3926

Qy 241 VTTSLGAGSYLALPALTHTHELRLDVEFKELADPGVLLFSGKSGPVEDFVSLAMVG 300
 |||||
 Db 3927 VTTSLGAGSYLALPALTHTHELRLDVEFKELADPGVLLFSGKSGPVEDFVSLAMVG 3986

Qy 301 GHLEFRYELGSLAVLRASAPLALGRWHRVSAERLNKDSLRVNGRFLVLRSPGKSGQL 360
 |||||
 Db 3987 GHLEFRYELGSLAVLRASAPLALGRWHRVSAERLNKDSLRVNGRFLVLRSPGKSGQL 4046

Qy 361 NLHTLLYLGVEPVPVLSPATNMSAHRFCVGEVSNKRLDLYTFLSQIGQCYDSS 420
 |||||
 Db 4047 NLHTLLYLGVEPVPVLSPATNMSAHRFCVGEVSNKRLDLYTFLSQIGQCYDSS 4106

Qy 421 PCRQPCQHGATCMRAGEYFQCLCRDGFKGLDCEHENPCQLRPEFCLHGTCQTRCLC 480
 |||||
 Db 4107 PCRQPCQHGATCMRAGEYFQCLCRDGFKGLDCEHENPCQLRPEFCLHGTCQTRCLC 4166

Qy 481 LRFSGPRCOQSGHGIABSDWHLSESGNDAPGQYGFHDDGFLAFPHFVFSRSLPEV 540
 |||||
 Db 4167 LRFSGPRCOQSGHGIABSDWHLSESGNDAPGQYGFHDDGFLAFPHFVFSRSLPEV 4226

Qy 541 PETIELEVRTSTASGILLWQVGEAGCKDFISLGLDGHLYFRYQLSGSEARLVSESD 600
 |||||
 Db 4227 PETIELEVRTSTASGILLWQVGEAGCKDFISLGLDGHLYFRYQLSGSEARLVSESD 4286

Qy 601 PINDGEHRVATLRREGRRGSIQVDEGEELVSGRSPGNVAVNAKSGVYGGAPDVATLGG 660
 |||||
 Db 4287 PINDGEHRVATLRREGRRGSIQVDEGEELVSGRSPGNVAVNAKSGVYGGAPDVATLGG 4346

Qy 661 RFSSGITGVKLVLSHARPGAPPQPLDLQHRAGAGANTRPCCPS 705
 |||||
 Db 4347 RFSSGITGVKLVLSHARPGAPPQPLDLQHRAGAGANTRPCCPS 4391

RESULT 2
 S18252
 heparan sulfate proteoglycan - mouse
 N:Alternate names: perlecan
 C:Species: Mus musculus (house mouse)
 C>Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text_change 05-Nov-1999
 C:Accession: S18252; A31917; B31917; S66460
 R:Noonan, D.M.; Fuller, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaaki, M.; Yamada, Y.; Hara, J. Biol. Chem. 266, 22939-22947, 1991
 A:Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteoglycan: adhesion molecule.
 A:Reference number: S18252; MUID:92078153; PMID:1744087
 A:Accession: S18252
 A:Molecule type: mRNA
 A:Residues: 1-3707 <NOO>
 A:Cross-references: EMBL:M77174; NID:G200295; PIDN:AAA39911.1; PID:G200296
 R:Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogeli, G.; Sasaki, M.; Yamada, Y.; Hara, J. Biol. Chem. 263, 16379-16387, 1998
 A:Title: Identification of cDNA clones encoding different domains of the basement membrane heparan sulfate proteoglycan.
 A:Reference number: A32680; MUID:89034110; PMID:2972708
 A:Accession: A31917
 A:Molecule type: mRNA
 A:Residues: 940-1601 <NO2>
 A:Cross-references: GB:J04054; NID:G200252; PIDN:AAA39899.1; PID:G200253
 A:Accession: B31917
 A:Molecule type: mRNA
 A:Residues: 1870-2600 <NO3>
 A:Cross-references: GB:J04055; NID:G200300; PIDN:AAA39912.1; PID:G200301
 R:Schulze, B.; Mann, K.; Battistutta, R.; Wiedemann, H.; Timpl, R.
 Eur. J. Biochem. 231, 551-556, 1995
 A:Title: Structural properties of recombinant domain III-3 of perlecan containing a globular domain.
 A:Reference number: S66460; MUID:95377282; PMID:7649154
 A:Accession: S66460
 A:Molecule type: protein
 A:Residues: 1272-1274,'X',1276,'X',1278-1279 <SCH>
 A:Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repeat
 C:Keywords: glycoprotein
 F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F:764-811/Domain: laminin-type EGF-like homology <LEG>
 F:1159-1206/Domain: laminin-type EGF-like homology <LEG7>
 F:1563-1610/Domain: laminin-type EGF-like homology <LEG8>
 F:1613-1668/Domain: laminin-type EGF-like homology <EG7>
 F:3163-3198/Domain: laminin-type EGF-like homology <EG8>
 F:3270-3423/Domain: laminin G repeat homology <LG2>
 F:464-3492/Domain: EGF homology <EGF7>
 F:1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.2%; Score 44; DB 2; Length 3707;
 Best Local Similarity 100.0%; Pred. No. 2,2e-36;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 SPTNLANKRQDFISGLVGGPEPRFDAGSGMATIRHTPLALG 69
Db 3030 SPTNLANKRQDFISGLVGGPEPRFDAGSGMATIRHTPLALG 3073

RESULT 3

A48825
Notch homolog Notch protein - mouse (fragment)
C:Species: Mus musculus (house mouse)

C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Sep-2002
C:Accession: A48825

R:Reaume, A.G.; Conlon, R.A.; Zirngibl, R.; Yamaguchi, T.P.; Rossant, J.
Dev. Biol. 154, 377-387, 1992

A:Title: Expression analysis of a Notch homologue in the mouse embryo.
A:Reference number: A48825; MUID:93050801; PMID:1426644

A:Accession: A48825
A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA
A:Residues: 1-861 <REA>

A:Experimental source: embryo
A>Note: sequence extracted from NCBI backbone (NCBIP:119144)

C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
F:26-57/Domain: EGF homology <EGF>

F:64-95/Domain: EGF homology <EGX1>
F:198-229/Domain: EGF homology <EGF2>

F:441-472/Domain: EGF homology <EGX2>

Query Match 1.3%; Score 9; DB 2; Length 861;

Best Local Similarity 100.0%; Pred. No. 2.5;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGTCQ 474

Db 240 PCLHGTCQ 248

RESULT 4

AGRT

agrin - rat

C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000

C:Accession: JH0399; A38856
R:Rupp, F.; Pavan, D.G.; Magill-Solc, C.; Cowan, D.M.; Scheller, R.H.

Neuron 6, 811-823, 1991
A:Title: Structure and expression of a rat agrin.

A:Reference number: JH0399; MUID:91222570; PMID:1851019

A:Accession: A38856
A:Molecule type: mRNA

A:Residues: 1-1779;1799-1959 <RUP>
A:Cross-references: GB:M64780; NID:9202798; PIDN:AAA40703.1; PID:9202800

A:Experimental source: embryonic spinal cord
A>Note: it is uncertain whether Met-1, Met-18, or Met-24 is the initiator

R:Rupp, F.; Oezcelik, T.; Linnak, M.; Peterson, K.; Francke, U.; Scheller, R.
J. Neurosci. 12, 3535-3544, 1992

A:Title: Structure and chromosomal localization of the mammalian agrin gene.
A:Reference number: A38856; MUID:92407628; PMID:13266508

A:Accession: A38856
A:Molecule type: mRNA

A:Residues: 1780-1798 <RU2>
A:Cross-references: GB:S44194

C:Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine 1
C:Comment: 90% of rat embryonic transcripts encode the variant labeled below as form 3.

Ylcholine receptor clustering activity.
C:Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea

C:Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
F:1-1959/Product: agrin, form 1 #status predicted <AG1>

F:1-1787,1799-1959/Product: agrin, form 4 #status predicted <AG4>
F:1-1779,1788-1959/Product: agrin, form 3 #status predicted <AG3>

F:1-1779,1788-1959/Product: agrin, form 5 #status predicted <AG5>
F:1-1143,1153-1959/Product: agrin, form 2 #status predicted <AG2>

F:22-50/Region: hydrophobic

F:88-137/Domain: Kazal proteinase inhibitor homology <KPI1>

F:163-212/Domain: Kazal proteinase inhibitor homology <KPI2>

F:236-284/Domain: Kazal proteinase inhibitor homology <KPI3>

F:307-356/Domain: Kazal proteinase inhibitor homology <KPI4>

F:381-429/Domain: Kazal proteinase inhibitor homology <KPI5>

F:446-494/Domain: Kazal proteinase inhibitor homology <KPI6>

F:511-559/Domain: Kazal proteinase inhibitor homology <KPI7>

F:540-542/Region: motor neuron attachment (L-R-E) motif

F:596-645/Domain: Kazal proteinase inhibitor homology <KPI8>

F:688-739/Domain: laminin-type EGF-like homology <LE1>

F:742-786/Domain: laminin-type EGF-like homology <LE2>

F:814-864/Domain: Kazal proteinase inhibitor homology <KPI9>

F:869-992/Region: serine/threonine-rich

F:1084-1086/Region: motor neuron attachment (L-R-E) motif

F:1147-1215/Region: serine/threonine-rich

F:1224-1257/Domain: EGF homology <EG1>

F:1287-1442/Domain: laminin G repeat homology <LG1>

F:1444-1476/Domain: EGF homology <EG2>

F:1483-1515/Domain: EGF homology <EG3>

F:1555-1706/Domain: laminin G repeat homology <LG2>

F:1713-1747/Domain: EGF homology <EG4>

F:1807-1959/Domain: laminin G repeat homology <LG3>

F:187-116,105-137,171-191,180-212,244-263,252-284,316-335,324-356,389-408,397-429,454-473,

-1476,1483-4494,1488-1504,1506-1515/Disulfide bonds: #status predicted

F:145,672,827,957/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.3%; Score 9; DB 1; Length 1959;

Best Local Similarity 100.0%; Pred. No. 5.4;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGTCQ 474

Db 1228 PCLHGTCQ 1236

RESULT 5

S18188

notch protein homolog - rat

C:Species: Rattus norvegicus (Norway rat)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002

C:Accession: S18188
R:Weinmaster, G.; Roberts, V.J.; Lemke, G.

Development 113, 199-205, 1991
A:Title: A homolog of Drosophila Notch expressed during mammalian development.

A:Reference number: S18188; MUID:92111383; PMID:1764995

A:Accession: S18188
A:Molecule type: mRNA

A:Residues: 1-2531 <WEI>
A:Cross-references: EMBL:X57405; NID:957634; PID:957635

C:Superfamily: notch protein; ankyrin repeat homology; EGF homology

F:987-1018/Domain: EGF homology <EGF1>

F:1025-1056/Domain: EGF homology <EGF2>

F:1233-1264/Domain: EGF homology <EGF3>

F:1917-1949/Domain: ankyrin repeat homology <AN1>

F:1950-1982/Domain: ankyrin repeat homology <AN2>

F:1984-2016/Domain: ankyrin repeat homology <AN3>

F:2017-2049/Domain: ankyrin repeat homology <AN4>

F:2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 1.3%; Score 9; DB 2; Length 2531;

Best Local Similarity 100.0%; Pred. No. 6.8;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGTCQ 474

Db 1029 PCLHGTCQ 1037

RESULT 6

A46019

notch-1 protein - mouse

N:Alternate names: notch protein

C:Species: Mus musculus (house mouse)

C;Date: 22-Sep-1993 #sequence revision 18-Nov-1994 #text change 07-Mar-2003
 C;Accession: A46019; S25144; C49175; B46438; A46438; PH1569; S32109
 R;del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkkins, N.A.; Copeland, N.G.; Grid
 Genomics 15, 259-264, 1993
 A;Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of
 A;Reference number: A46019; MUID:93194170; PMID:8449489
 A;Accession: A46019
 A;Status: not compared with conceptual translation
 A;Molecule type: nucleic acid
 A;Residues: 1-2531
 A;Cross-references: GB:Z11886; GB:847228; NID:9288502; PIDN:CAA77941.1; PID:G288503
 A;Note: sequence extracted from NCBI backbone (NCBIP:1273118)
 R;Franco del Amo, F.; Smith, D.B.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.;
 submitted to the EMBL Data Library, April 1992
 A;Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, suggest
 A;Reference number: S25144
 A;Accession: S25144
 A;Molecule type: mRNA
 A;Residues: 1551-2108, 'O', 2110-2114, 'ALP', 2118-2170 <FRA>
 R;Lardelli, M.; Lendahl, U.
 Exp. Cell Res. 204, 364-372, 1993
 A;Title: Moch A and Moch B--two mouse Notch homologues coexpressed in a wide variety of
 A;Reference number: A49175; MUID:93178563; PMID:8440332
 A;Accession: C49175
 A;Status: preliminary; nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1161-1547 <LAR>
 A;Cross-references: EMBL:X68278; NID:9287987; PIDN:CAA48339.1; PID:G287988
 A;Experimental source: embryo
 A;Note: sequence extracted from NCBI backbone (NCBIP:126159)
 R;Kopan, R.; Weintraub, H.
 J. Cell Biol. 121, 631-641, 1993
 A;Title: Mouse notch: expression in hair follicles correlates with cell fate determinati
 A;Reference number: A46438; MUID:93252998; PMID:8486742
 A;Accession: B46438
 A;Status: preliminary
 A;Molecule type: nucleic acid
 A;Residues: 1865-1932, 'RR', 1935-1937, 'L', 1938-1967, 'I', 1969-2044, 'IE', 2047-2052, 'S', 2054
 A;Experimental source: embryo
 A;Note: sequence extracted from NCBI backbone (NCBIN:131246, NCBIP:131247)
 C;Comment: This protein has many EGF repeats and lin-12[1172]/Notch repeats.
 C;Comment: This protein is one of the neurogenic proteins controlling the decision betwe
 C;Genetics:
 A;Gene: notch-1
 A;Map position: 2
 A;Note: proximal region of chromosome 2
 C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F;106-138/Domain: EGF homology <EGF1>
 F;144-175/Domain: EGF homology <EGF1>
 F;222-254/Domain: EGF homology <EGF2>
 F;261-292/Domain: EGF homology <EGF2>
 F;339-370/Domain: EGF homology <EGF3>
 F;416-449/Domain: EGF homology <EGF3>
 F;456-487/Domain: EGF homology <EGF4>
 F;494-525/Domain: EGF homology <EGF4>
 F;532-563/Domain: EGF homology <EGF5>
 F;607-638/Domain: EGF homology <EGF6>
 F;682-713/Domain: EGF homology <EGF7>
 F;757-788/Domain: EGF homology <EGF8>
 F;795-826/Domain: EGF homology <EGF9>
 F;870-904/Domain: EGF homology <EGF10>
 F;911-942/Domain: EGF homology <EGF11>
 F;949-980/Domain: EGF homology <EGF12>
 F;987-1018/Domain: EGF homology <EGF13>
 F;1025-1056/Domain: EGF homology <EGF14>
 F;1063-1094/Domain: EGF homology <EGF15>
 F;1149-1180/Domain: EGF homology <EGF16>
 F;1187-1218/Domain: EGF homology <EGF17>
 F;1233-1264/Domain: EGF homology <EGF18>
 F;1352-1383/Domain: EGF homology <EGF19>
 F;1391-1425/Domain: EGF homology <EGF20>
 F;1917-1948/Domain: ankyrin repeat homology <ANI>

F;1949-1991/Domain: ankyrin repeat homology <AN2>
 F;1983-2015/Domain: ankyrin repeat homology <AN3>
 F;2016-2048/Domain: ankyrin repeat homology <AN4>
 F;2049-2081/Domain: ankyrin repeat homology <AN5>
 Query Match 1.3%; Score 9; DB 2; Length 2531;
 Best Local Similarity 100.0%; Pred. No. 6.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 466 PCLHGGTCC 474
 Db 1029 PCLHGGTCC 1037
 RESULT 7
 T48789
 TFS1 related protein [imported] - Neurospora crassa
 N;Alternate names: protein 13E11.370
 C;Species: Neurospora crassa
 C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
 C;Accession: T48789
 R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
 submitted to the Protein Sequence Database, April 2000
 A;Reference number: Z24541
 A;Accession: T48789
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-215 <SCH>
 A;Cross-references: EMBL:AL333820; GSPDB:GN00112; NCSP:13E11.370
 A;Experimental source: cosmid contig 13E11, strain 74
 C;Genetics:
 A;Gene: NCSP:13E11.370
 A;Map position: 2
 Query Match 1.1%; Score 8; DB 2; Length 215;
 Best Local Similarity 100.0%; Pred. No. 7.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 90 LAPVNGTS 97
 Db 136 LAPVNGTS 143
 RESULT 8
 A70772
 Probable sulfate adenylyate transferase subunit 2 - Mycobacterium tuberculosis (strain H3
 C;Species: Mycobacterium tuberculosis
 C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C;Accession: A70772
 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A;Reference number: A70500; MUID:98295987; PMID:9634230
 A;Accession: A70772
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-332 <COL>
 A;Cross-references: GB:Z73419; GB:AL123456; NID:93261573; PIDN:CAA97751.1; PID:ig1322409
 A;Experimental source: strain H37Rv
 C;Genetics:
 A;Gene: cysB
 A;Superfamily: nodulation protein nodP
 Query Match 1.1%; Score 8; DB 2; Length 332;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 279 VLLFSGGK 286
 Db 64 VLLFSGGK 71

RESULT 9
 A95935
 conserved hypothetical membrane protein SMB21241 [imported] - Sinorhizobium meliloti (sc
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: A95935
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 C:Accession: A95935
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-430 <KUR>
 A:Cross-references: GB:AL591985; PIDN:CAC49145.1; PID:G15140630; GSPDB:GN00167
 A:Experimental source: strain 1021 megaplasmid pSymB
 R:Galibert, P.; Finan, T.M.; Long, S.R.; Fuhrer, A.; Abola, P.; Ampe, P.; Barloy-Hubler,
 Peia, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hehault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: SMB21241
 A:Genome: plasmid

Query Match 1.1%; Score 8; DB 2; Length 430;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 SYLALPAL 258
 |||||
 DB 208 SYLALPAL 215

RESULT 10
 AH2015
 sodium/solute symporter [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AH2015
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shampo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AH2015
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-497 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA878044.1; PID:G17135498; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all1678

Query Match 1.1%; Score 8; DB 2; Length 497;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 STASGLLL 558
 |||||
 DB 325 STASGLLL 332

RESULT 11
 F70012
 leucyl aminopeptidase homolog yuiE - Bacillus subtilis

C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 03-Mar-2003
 C:Accession: F70012
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Brouillet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
 A.; Ehrlich, S.D.; Emmsarson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A.; Muthers, S.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Y, M.; Ogawa, K.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
 Rieger, M.; Rivolta, C.; Schlotter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
 A.; Authors: Schleich, S.; Schroeter, P.; Terstra, P.; Tognoni, A.; Tsao, V.; Uchiyama
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: F70012
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1-500 <KUN>
 A:Cross-references: GB:Z99120; GB:AL009126; NID:G2635613; PIDN:CABL5195.1; PID:G2635702
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yuiE
 C:Superfamily: Cytosol aminopeptidase

Query Match 1.1%; Score 8; DB 2; Length 500;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 653 DVATLTGG 660
 |||||
 DB 367 DVATLTGG 374

RESULT 12
 T00732
 hypothetical protein F22013.27 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999
 C:Accession: T00732
 R:Shinn, P.; Bushler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Con
 sologs, A.; Ecker, J.R.
 submitted to the EMBL Data Library, April 1998
 A:Description: Genomic sequence for Arabidopsis thaliana BAC F22013.
 A:Reference number: Z14200
 A:Accession: T00732
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-748 <SHI>
 A:Cross-references: EMBL:AC003981; NID:G3063438; PID:G3063465; GSPDB:GN00059; ATSP:F2201
 C:Genetics:
 A:Gene: ATSP:F22013.27
 A:Map position: 1
 A:Introns: 329/3; 429/3; 460/2

Query Match 1.1%; Score 8; DB 2; Length 748;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 394 VSVNGKRL 401
 |||||
 DB 403 VSVNGKRL 410

RESULT 13
 T39081
 hypothetical protein SPACT4.03c - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000

C:Accession: T39081
 R:Gentles, S.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, September 1997.
 A:Reference number: Z11826
 A:Accession: T39081
 A>Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-886 <GEN>
 A:Cross-references: EMBL:Z99532; PIDN:CAB16720.1; GSPDB:GN00066; SPDB:SPACTD4.03C
 A:Experimental source: strain 972h; cosmid c7D4
 C:Genetics:
 A:Gene: SPDB:SPACTD4.03C
 A:Map position: 1
 C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPACTD4.03C

Query Match 1.1%; Score 8; DB 2; Length 886;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 394 VSVNGKRL 401
 |||||
 Db 828 VSVNGKRL 835

RESULT 14

A49175
 Motch B protein - mouse (fragment)
 N:Alternate names: Notch homolog
 C:Species: Mus musculus (house mouse)
 C:Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 08-Sep-2002
 C:Accession: A49175; PH1570; S32113
 R:Lardelli, M.; Lendahl, U.
 Exp. Cell Res. 204, 364-372, 1993
 A:Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety of tissues
 A:Reference number: A49175; MUID:93178563; PMID:8440332
 A:Accession: A49175
 A>Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-1203 <LAR>
 A:Cross-references: EMBL:X68279; NID:9287989; PIDN:CAA48340.1; PID:G287990
 A:Experimental source: embryo
 A:Note: sequence extracted from NCBI backbone (NCBIP:126158)
 C:Comment: This protein has many EGF repeats and lin-12/Notch repeats.
 C:Comment: This protein is one of the neurogenic proteins controlling the decision between neurogenesis and proliferation.
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:143-174/Domain: EGF homology <EGX1>
 F:482-513/Domain: EGF homology <EGF1>
 F:560-591/Domain: EGF homology <EGF>
 F:674-705/Domain: EGF homology <EGX2>
 F:712-743/Domain: EGF homology <EGF3>
 F:836-867/Domain: EGF homology <EGX3>

Query Match 1.1%; Score 8; DB 2; Length 1203;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 426 PCQHGATC 433
 |||||
 Db 840 PCQHGATC 847

RESULT 15

S28764
 neurocan precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 04-Feb-2000
 C:Accession: S28764
 R:Rauch, U.; Karthikeyan, L.; Maurel, P.; Margolis, R.U.; Margolis, R.K.
 J. Biol. Chem. 267, 19536-19547, 1992
 A:Title: Cloning and primary structure of neurocan, a developmentally regulated, aggregated proteoglycan core protein
 A:Reference number: S28764; MUID:92406907; PMID:1326557
 A:Accession: S28764
 A:Molecule type: mRNA

A:Residues: 1-1257 <RAU>
 A:Cross-references: EMBL:M97161; NID:G205649; PIDN:AAC37679.1; PID:G205650
 C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EG
 C:Keywords: chondroitin sulfate proteoglycan; glycoprotein
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-1257/Product: neurocan #status predicted <MAT>
 F:176-253/Domain: link protein repeat homology <LNKL>
 F:274-355/Domain: link protein repeat homology <MAT>
 F:364-366/Region: cell attachment (R-G-D) motif
 F:953-984/Domain: EGF homology <EGF>
 F:1029-1149/Domain: C-type lectin homology <LCH>
 F:1156-1212/Domain: complement factor H repeat homology <FHD>
 F:121.339,737,967,1164/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:372.410/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
 F:944/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental

Query Match 1.1%; Score 8; DB 2; Length 1257;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 466 PCLHGATC 473
 |||||
 Db 957 PCLHGATC 964

Search completed: March 9, 2004, 17:31:26
 Job time : 26.1076 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:21:48 ; Search time 17.0517 Seconds
(without alignments)
2152.832 Million cell updates/sec

Title: US-10-006-011a-3
Perfect score: 705
Sequence: 1 EIKITFRDSDAGMLLYNQ.....QLDLQHRQAQANTRPCPS 705

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Table with columns: Result No., Score, Query Match, Length, DB ID, Description. Contains 33 entries of protein search results.

Table with columns: ID, PGEM HUMAN, STANDARD, PRT; 4391 AA. Lists various protein sequences and their identifiers.

ALIGNMENTS

RESULT 1
PGEM_HUMAN
ID PG8160; Q16287; Q9H3V5; PRT; 4391 AA.
AC 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basemat membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (perlecan) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. PubMed=1730769;
RX MEDLINE=92112994; PubMed=1730769;
RA Kallunki P., Tryggvason K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor.";
RL J. Cell Biol. 116:559-571(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Colon, and Skin;
RC MEDLINE=92235084; PubMed=1569102;
RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG2/perlecan) A chimeric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor.";
RL J. Biol. Chem. 267:8544-8557(1992).
RN [3]
RP SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TYR-1532.
RX MEDLINE=20553141; PubMed=11101850;
RA Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,
RA Beighton P., Ben-Hamida C., Hammouda H., Craud C., White P.S.,
RA Samson D., Urtizbera J.A., Lehmann-Horn F., Weissenbach J.,
RA Hentati F., Fontaine B.;
RT "Perlecan, the major proteoglycan of basement membranes, is altered in
RT patients with Schwartt-Jampel syndrome (chondrodystrophic myotonia).";
RL Nat. Genet. 26:480-483(2000).
RN [4]
RP SEQUENCE OF 1016-1470 FROM N.A.
RX TISSUE=Colon;
RC MEDLINE=91365376; PubMed=1679749;
RA Dodge G.R., Kovaliszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA Yi H.F., Iozzo R.V.;
RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1.";
RL Genomics 10:673-680(1991).
RN [5]
RP SEQUENCE OF 950-1396 FROM N.A.
RX TISSUE=Fibrosarcoma;

RX MEDLINE=92120660; PubMed=1685141;
RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA Trygvsason K.;
RT Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPC2) to lp36.1--p35 and identification of
RT a BamHI restriction fragment length polymorphism.;
RL Genomics 11:389-396(1991).
RN [6]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=94052171; PubMed=8234307;
RA Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT Structural characterization of the complete human perlecan gene and
RT its promoter.;
RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN [7]
RP CARBOHYDRATE-LINKAGE SITE ASN-2121.
RX MEDLINE=22660472; PubMed=12754519;
RA Zhang H., Li X.-J., Martin D.B., Aebbersold R.;
RT Identification and quantification of N-linked glycoproteins using
RT hydrazide chemistry, stable isotope labeling and mass spectrometry.;
RL Nat. Biotechnol. 21:660-666(2003).
CC !- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC !- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC !- SUBCELLULAR LOCATION: Extracellular.
CC !- TISSUE SPECIFICITY: Found in the basement membranes.
CC !- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC !- DISEASE: Defects in HSPC2 are the cause of Schwartz-Jampel
CC syndrome (SJS1) (MIM:255800); a rare autosomal recessive disorder
CC characterized by permanent myotonia (prolonged failure of muscle
CC relaxation) and skeletal dysplasia, resulting in reduced stature,
CC kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
CC !- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC !- SIMILARITY: Contains 11 laminin EGF-like domains.
CC !- SIMILARITY: Contains 3 laminin IV domains.
CC !- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
CC !- SIMILARITY: Contains 3 laminin G-like domains.
CC !- SIMILARITY: Contains 4 EGF-like domains.
CC !- SIMILARITY: Contains 1 SEA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sb-sib.ch).
CC
CC EMBL; X62515; CRA44373.1; -
CC EMBL; M85289; ABA52700.1; -
CC EMBL; AL445795; CAC18534.1; -
CC EMBL; M64283; ARA52699.1; -
CC EMBL; S76436; AAB21121.2; -
CC EMBL; L22078; -; NOT_ANNOTATED_CDS.
CC PIR; A38096; A38096.
CC HSPG; P00740; 1RDM.
CC Siena-2DPAGE; P98160; -
CC Genew; HGNC:5273; HSPG2.
CC MIM; 142461; -
CC MIM; 255800; -
CC InterPro; IPR008985; Cora_like_lec_gl.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR006209; EGF_like.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003598; Ig_c2.

DR InterPro; IPR003596; Ig_v.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF00008; EGF_4.
DR Pfam; PF00047; ig_22.
DR Pfam; PF00052; laminin_B_3.
DR Pfam; PF00053; laminin_EGF_7.
DR Pfam; PF00054; laminin_G_3.
DR Pfam; PF00057; ldl_recept_a_4.
DR Pfam; PF01390; SEA_1.
DR PRINTS; PR00261; LDLRECEPTOR.
DR ProDom; PD003031; Laminin_B_3.
DR SMART; SM00181; EGF_15.
DR SMART; SM00180; EGF_Lam_12.
DR SMART; SM00409; IG_22.
DR SMART; SM00408; IGC2_21.
DR SMART; SM00406; IGV_7.
DR SMART; SM00281; Lamb_3.
DR SMART; SM00282; LamG_3.
DR SMART; SM00192; LDLa_4.
DR SMART; SM0200; SEA_1.
DR PROSITE; PS00022; EGF_1_9.
DR PROSITE; PS01186; EGF_2_6.
DR PROSITE; PS00026; EGF_3_4.
DR PROSITE; PS50835; IG_LIKE_22.
DR PROSITE; PS00025; LAM_G_DOMAIN_3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF_11.
DR PROSITE; PS01209; LDLa_1_4.
DR PROSITE; PS50068; LDLa_2_4.
DR PROSITE; PS50024; SEA_1.
KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW Extracellular matrix; EGF-like domain; Disease mutation.
FT SIGNAL 1 21
FT CHAIN 22 4391
FT DOMAIN 80 194
FT DOMAIN 198 235
FT DOMAIN 284 320
FT DOMAIN 334 360
FT DOMAIN 367 404
FT DOMAIN 405 504
FT DOMAIN 521 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1955
FT DOMAIN 1956 2051
FT DOMAIN 2052 2151
FT DOMAIN 2152 2244
FT DOMAIN 2245 2340
FT DOMAIN 2341 2437
FT DOMAIN 2438 2533
FT DOMAIN 2534 2629
FT DOMAIN 2630 2726

LDL-RECEPTOR CLASS A 1.
LDL-RECEPTOR CLASS A 2.
LDL-RECEPTOR CLASS A 3.
LDL-RECEPTOR CLASS A 4.
IG-LIKE C2-TYPE 1.
LAMININ EGF-LIKE 1 (N-TERMINAL).
LAMININ DOMAIN IV 1 (DOMAIN III A).
LAMININ EGF-LIKE 1 (C-TERMINAL).
LAMININ EGF-LIKE 2.
LAMININ EGF-LIKE 3.
LAMININ EGF-LIKE 4 (INCOMPLETE).
LAMININ EGF-LIKE 5 (N-TERMINAL).
LAMININ DOMAIN IV 2 (DOMAIN III B).
LAMININ EGF-LIKE 5 (C-TERMINAL).
LAMININ EGF-LIKE 6.
LAMININ EGF-LIKE 7.
LAMININ EGF-LIKE 8.
LAMININ EGF-LIKE 9 (N-TERMINAL).
LAMININ DOMAIN IV 3 (DOMAIN III C).
LAMININ EGF-LIKE 9 (C-TERMINAL).
LAMININ EGF-LIKE 10.
LAMININ EGF-LIKE 11.
IG-LIKE C2-TYPE 2.
IG-LIKE C2-TYPE 3.
IG-LIKE C2-TYPE 4.
IG-LIKE C2-TYPE 5.
IG-LIKE C2-TYPE 6.
IG-LIKE C2-TYPE 7.
IG-LIKE C2-TYPE 8.
IG-LIKE C2-TYPE 9.
IG-LIKE C2-TYPE 10.
IG-LIKE C2-TYPE 11.
IG-LIKE C2-TYPE 12.

Qy	Db	Query Match	100.0%	Score	705	DB	1	Length	4391
Matches	705	Conservative	0	Mismatches	0	Indels	0	Gaps	0
Qy	1	EIKITFRPDSADGMLLYNGQKRVPGSPTNLNANRQDPFISFGLVGGPFRFRFDAGSGMAYI	60						
Db	3687	EIKITFRPDSADGMLLYNGQKRVPGSPTNLNANRQDPFISFGLVGGPFRFRFDAGSGMAYI	3746						
Qy	61	RHPTPLALGHFHTVTLRSITOGSLIVGLDAPVNGTSQKFGQLDNEELYLGGYDYGA	120						
Db	3747	RHPTPLALGHFHTVTLRSITOGSLIVGLDAPVNGTSQKFGQLDNEELYLGGYDYGA	3806						
Qy	121	IPKAGLSSGFIGVRELIQGEIVFHDNLNLTAGHSHCPTCRDRPCQNGOCHDSESS	180						
Db	3807	IPKAGLSSGFIGVRELIQGEIVFHDNLNLTAGHSHCPTCRDRPCQNGOCHDSESS	3866						
Qy	181	YVCVCFAGTGSCEHSQALHCHPEACGPDATCVNRPDGRYTCRHLGRSGLRCEEGVT	240						
Db	3867	YVCVCFAGTGSCEHSQALHCHPEACGPDATCVNRPDGRYTCRHLGRSGLRCEEGVT	3926						
Qy	241	VTFPSLSGAGSYLALPALNTNHHLELDVFEFKPLADGVLLFSGGKSGPVEDFVSLAMVG	300						
Db	3927	VTFPSLSGAGSYLALPALNTNHHLELDVFEFKPLADGVLLFSGGKSGPVEDFVSLAMVG	3986						
Qy	301	GHLEFRYELGSLAVLRSAPPLALGRWHRVSAERLNKDGSLRVNGRVPVLRSSPGKSOGL	360						
Db	3987	GHLEFRYELGSLAVLRSAPPLALGRWHRVSAERLNKDGSLRVNGRVPVLRSSPGKSOGL	4046						
Qy	361	NLHTLLYLGVEVPSVPLSPATNMSAHRFCVGVSVNGKRLDLYFLGSGIGQCYDSS	420						
Db	4047	NLHTLLYLGVEVPSVPLSPATNMSAHRFCVGVSVNGKRLDLYFLGSGIGQCYDSS	4106						
Qy	421	PCBOPCOHGATCMPAGEYFQCLCRDGFKGLDCEHEENPCQLRPFCLHGTCQGTCLC	480						
Db	4107	PCBOPCOHGATCMPAGEYFQCLCRDGFKGLDCEHEENPCQLRPFCLHGTCQGTCLC	4166						
Qy	481	LPFGSPRCQGGHGIASDWHLEGGNDAPQGYAFHDDGLAPGHVFSRSLPEV	540						
Db	4167	LPFGSPRCQGGHGIASDWHLEGGNDAPQGYAFHDDGLAPGHVFSRSLPEV	4226						
Qy	541	PETIELEVRTSTASGLLLWQGVVEGAGQKDFISLGLQDGHVFRYQLGSGEARLVSED	600						
Db	4227	PETIELEVRTSTASGLLLWQGVVEGAGQKDFISLGLQDGHVFRYQLGSGEARLVSED	4286						
Qy	601	PINDGEMHRTALREGRGSIQVDBELVSGRSPGNVAVNAKGSVYTGADPVATLTGG	660						
Db	4287	PINDGEMHRTALREGRGSIQVDBELVSGRSPGNVAVNAKGSVYTGADPVATLTGG	4346						
Qy	661	RFSSGITGCVKMLVLSARPAPPPOPLDQHRAGANTRPCPS	705						
Db	4347	RFSSGITGCVKMLVLSARPAPPPOPLDQHRAGANTRPCPS	4391						

RESULT 2
 PGEM_MOUSE STANDARD; PRT; 3707 AA.
 AC Q05753;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Basement membrane-specific heparan sulfate proteoglycan core protein precursor (HSPG) (perlecan) (PLC).
 GN HSPG2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 [1]
 RN Noonan D.M., Fullie A., Valente P., Cai S., Horigan E., Sasaki M.,
 RC TISSUE=Melanoma; PubMed=1744087;
 RX MEDLINE=92078153;
 RA Noonan D.M., Hassell J.R.,

Yamada Y., Hassell J.R.;
 "The complete sequence of perlecan, a basement membrane heparan sulfate proteoglycan, reveals extensive similarity with laminin A chain, low density lipoprotein-receptor, and the neural cell adhesion molecule";
 J. Biol. Chem. 266:22939-22947(1991).
 [2]
 SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE. MEDLINE=89034110; PubMed=2972708;
 Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M., Yamada Y., Hassell J.R.;
 "Identification of cDNA clones encoding different domains of the basement membrane heparan sulfate proteoglycan";
 J. Biol. Chem. 263:16379-16387(1988).
 CC -!- FUNCTION: This protein is an integral component of basement membranes. It is responsible for the fixed negative electrostatic charge and is involved in the charge-selective ultrafiltration properties. It serves as an attachment substrate for cells.
 CC -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in dimers or stellate structures. It interacts with other basement membrane components such as laminin, prolargin and collagen type IV.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: Found in the basement membranes.
 CC -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
 CC -!- SIMILARITY: Contains 11 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 3 laminin IV domains.
 CC -!- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
 CC -!- SIMILARITY: Contains 3 laminin G-like domains.
 CC -!- SIMILARITY: Contains 1 EGF-like domain.
 CC -!- SIMILARITY: Contains 1 SEA domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

 DR EMBL; M7174; AAA39911.1; -;
 DR EMBL; J04054; AAA39999.1; -;
 DR PIR; S18252; S18252.
 DR PDB; 1GL4; 28-NOV-01.
 DR MGD; MGI.96257; HSPG2.
 DR GO; GO:0005604; C:basement membrane; IDA.
 DR InterPro; IPR008985; ConA_like lec_gl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR02049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00047; ig; 15.
 DR Pfam; PF00052; laminin_B; 3.
 DR Pfam; PF00053; laminin_EGF; 7.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF00057; ldl_recept_a; 4.
 DR Pfam; PF01390; SEA; 1.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM003031; Laminin_B; 3.
 DR SMART; SM00180; EGF_Lam; 7.
 DR SMART; SM00408; IGC2; 14.
 DR SMART; SM00281; Lamb; 3.
 DR SMART; SM00282; Lamg; 3.

DR SMART; SMO0192; LDLa; 4.
DR SMART; SMO0200; SEA; 1.
DR PROSITE; PS00022; EGF_1; 8.
DR PROSITE; PS01186; EGF_2; 5.
DR PROSITE; PS00026; EGF_3; 4.
DR PROSITE; PS00835; IG LIKE; 15.
DR PROSITE; PS00025; LAM G DOMAIN; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR PROSITE; PS02009; LDLRA_1; 4.
DR PROSITE; PS02068; LDLRA_2; 4.
DR PROSITE; PS00024; SEA; 1.
DR Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW Extracellular matrix; EGF-like domain; 3D-structure.
FT SIGNAL 1 21
FT CHAIN 22 3707
FT DOMAIN 80 194
FT DOMAIN 195 234
FT DOMAIN 281 319
FT DOMAIN 320 359
FT DOMAIN 360 403
FT DOMAIN 404 504
FT DOMAIN 521 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1677 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1954
FT DOMAIN 1955 2049
FT DOMAIN 2050 2148
FT DOMAIN 2149 2244
FT DOMAIN 2245 2343
FT DOMAIN 2344 2436
FT DOMAIN 2437 2532
FT DOMAIN 2533 2619
FT DOMAIN 2620 2720
FT DOMAIN 2721 2809
FT DOMAIN 2810 2895
FT DOMAIN 2896 2980
FT DOMAIN 2984 3162
FT DOMAIN 3163 3241
FT DOMAIN 3245 3425
FT DOMAIN 3518 3705
FT SITE 65 67
FT SITE 71 73
FT SITE 76 78
FT SITE 3615 3617
FT DISULFID 199 212
FT DISULFID 206 225
FT DISULFID 219 234
FT DISULFID 285 297
FT DISULFID 292 310
FT DISULFID 304 319
FT DISULFID 325 337
FT DISULFID 332 350
FT DISULFID 344 359
FT DISULFID 368 381

DISULFID 375 394
DISULFID 388 403
DISULFID 428 479
DISULFID 764 773
DISULFID 766 780
DISULFID 793 792
DISULFID 795 811
DISULFID 814 829
DISULFID 816 839
DISULFID 842 851
DISULFID 854 869
DISULFID 1159 1168
DISULFID 1161 1175
DISULFID 1178 1187
DISULFID 1190 1206
DISULFID 1209 1224
DISULFID 1211 1234
DISULFID 1237 1246
DISULFID 1249 1263
DISULFID 1275 1287
DISULFID 1277 1293
DISULFID 1295 1304
DISULFID 1307 1322
DISULFID 1563 1572
DISULFID 1565 1579
DISULFID 1582 1591
DISULFID 1594 1610
DISULFID 1613 1628
DISULFID 1615 1638
DISULFID 1641 1650
DISULFID 1653 1668
DISULFID 1792 1839
DISULFID 1886 1932
DISULFID 1976 2021
DISULFID 2073 2118
DISULFID 2170 2215
DISULFID 2268 2313

Query Match 6.2%; Score 44; DB 1; Length 3707;
Best Local Similarity 100.0%; Pred.No. 4.6e-35;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 SPTNLANNRQPDFISFGLVGGRRPFRFDAGSGMATIRHPPTPLALG 69
|||||
Db 3030 SPTNLANNRQPDFISFGLVGGRRPFRFDAGSGMATIRHPPTPLALG 3073

RESULT 3
AGRI_RAT STANDARD; PRT; 1959 AA.
ID_AGRI_RAT Q63034;
AC P25304; Q63034;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Agrin precursor.
GN AGRN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
RC TISSUE=Embryonic spinal cord;
RX MEDLINE=91222570; PubMed=1851019;
RA Rupp F., Fayan D.G., Megill-Solc C., Cowan D.M., Scheller R.H.;
RT "Structure and expression of a rat agrin."
RL Neuron 6:811-823(1991).
RN [2]
RP SEQUENCE OF 1777-1801 FROM N.A.
RX MEDLINE=92407628; PubMed=1326608;
RA Rupp F., Oezcelik T., Linnal M., Peterson K., Francke U., Scheller R.;
RT "Structure and chromosomal localization of the mammalian agrin gene."
RL J. Neurosci. 12:3535-3544(1992).

```

-!- FUNCTION: Component of the basal lamina that causes the aggregation of acetylcholine receptors and acetylcholine-esterase on the surface of muscle fibers of the neuromuscular junction.
-!- SUBUNIT: Binds to laminin.
-!- CELLULAR LOCATION: Synaptic basal lamina at the neuromuscular junction.
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=5;
Comment=Additional isoforms seem to exist. Isoforms differ in their acetylcholine receptor clustering activity;
Name=1;
IsoId=P25304-1; Sequence=Displayed;
Name=2;
IsoId=P25304-2; Sequence=VSP_001365;
Name=3;
IsoId=P25304-3; Sequence=VSP_001366;
Name=4;
IsoId=P25304-4; Sequence=VSP_001367;
Name=5;
IsoId=P25304-5; Sequence=VSP_001368;
-!- TISSUE SPECIFICITY: Embryonic nervous system and muscle.
-!- DEVELOPMENTAL STAGE: More abundant early in development.
-!- PTM: Contains heparan sulfate chains as well as N-linked and O-linked oligosaccharides (By similarity).
-!- SIMILARITY: Contains 9 Kazal-like domains.
-!- SIMILARITY: Contains 2 laminin EGF-like domains.
-!- SIMILARITY: Contains 4 EGF-like domains.
-!- SIMILARITY: Contains 1 SEA domain.
-!- SIMILARITY: Contains 3 laminin G-like domains.
-!- CAUTION: It is uncertain whether Met-1, Met-18 or Met-24 is the initiator.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL; M64780; AAA40703.1; -
EMBL; M64780; AAA40702.1; ALT_INIT.
EMBL; S44194; AAB2326.1; -
PIR; JH0399; AGRT.
HSP; P00740; LEDM.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR003645; FOLN.
DR InterPro; IPR002350; kazal.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF00008; EGF_4.
DR Pfam; PF00053; laminin_EGF; 2.
DR Pfam; PF00554; laminin_G; 3.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00180; EGF_Lam; 2.
DR SMART; SM00274; FOLN; 8.
DR SMART; SM00280; KAZAL; 9.
DR SMART; SM00282; LamG; 3.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00026; EGF_3; 4.
DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 1959 AGRIN.

```

FT	DOMAIN	65	137	KAZAL-LIKE 1.
FT	DOMAIN	141	212	KAZAL-LIKE 2.
FT	DOMAIN	213	284	KAZAL-LIKE 3.
FT	DOMAIN	287	356	KAZAL-LIKE 4.
FT	DOMAIN	361	429	KAZAL-LIKE 5.
FT	DOMAIN	430	494	KAZAL-LIKE 6.
FT	DOMAIN	495	559	KAZAL-LIKE 7.
FT	DOMAIN	563	645	KAZAL-LIKE 8.
FT	DOMAIN	688	741	LAMININ EGF-LIKE 1.
FT	DOMAIN	742	788	LAMININ EGF-LIKE 2.
FT	DOMAIN	794	864	KAZAL-LIKE 9.
FT	DOMAIN	1023	1145	SEA.
FT	DOMAIN	1220	1258	EGF-LIKE 1.
FT	DOMAIN	1263	1439	LAMININ G-LIKE 1.
FT	DOMAIN	1440	1477	EGF-LIKE 2.
FT	DOMAIN	1479	1516	EGF-LIKE 3.
FT	DOMAIN	1526	1708	LAMININ G-LIKE 2.
FT	DOMAIN	1709	1748	EGF-LIKE 4.
FT	DOMAIN	1784	1956	LAMININ G-LIKE 3.
FT	DOMAIN	869	992	SER/THR-RICH.
FT	DOMAIN	1147	1215	POTENTIAL.
FT	DISULFID	97	116	POTENTIAL.
FT	DISULFID	105	137	POTENTIAL.
FT	DISULFID	171	191	POTENTIAL.
FT	DISULFID	180	212	POTENTIAL.
FT	DISULFID	244	263	POTENTIAL.
FT	DISULFID	252	284	POTENTIAL.
FT	DISULFID	316	335	POTENTIAL.
FT	DISULFID	324	356	POTENTIAL.
FT	DISULFID	389	408	POTENTIAL.
FT	DISULFID	397	429	POTENTIAL.
FT	DISULFID	454	473	POTENTIAL.
FT	DISULFID	462	494	POTENTIAL.
FT	DISULFID	518	538	POTENTIAL.
FT	DISULFID	527	559	POTENTIAL.
FT	DISULFID	604	624	POTENTIAL.
FT	DISULFID	613	645	POTENTIAL.
FT	DISULFID	688	700	POTENTIAL.
FT	DISULFID	690	707	BY SIMILARITY.
FT	DISULFID	709	718	BY SIMILARITY.
FT	DISULFID	721	739	BY SIMILARITY.
FT	DISULFID	742	754	BY SIMILARITY.
FT	DISULFID	744	761	BY SIMILARITY.
FT	DISULFID	763	772	BY SIMILARITY.
FT	DISULFID	775	786	BY SIMILARITY.
FT	DISULFID	823	843	POTENTIAL.
FT	DISULFID	832	864	POTENTIAL.
FT	DISULFID	1224	1235	BY SIMILARITY.
FT	DISULFID	1229	1246	BY SIMILARITY.
FT	DISULFID	1248	1257	BY SIMILARITY.
FT	DISULFID	1444	1455	POTENTIAL.
FT	DISULFID	1449	1465	POTENTIAL.
FT	DISULFID	1467	1476	POTENTIAL.
FT	DISULFID	1483	1494	BY SIMILARITY.
FT	DISULFID	1488	1504	BY SIMILARITY.
FT	DISULFID	1506	1515	BY SIMILARITY.
FT	DISULFID	1713	1727	BY SIMILARITY.
FT	DISULFID	1721	1736	BY SIMILARITY.
FT	DISULFID	1738	1747	BY SIMILARITY.
FT	CARBOHYD	145	145	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	672	672	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	827	827	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	957	957	N-LINKED (GLCNAC. . .)
FT	VARSPPLIC	1144	1152	Missing (in isoform 2).
FT	VARSPPLIC	1780	1798	Missing (in isoform 3).
FT	VARSPPLIC	1788	1798	/FTId=VSP_001366.
FT	VARSPPLIC	1780	1787	Missing (in isoform 4).
FT	VARSPPLIC	1780	1787	/FTId=VSP_001367.
FT	VARSPPLIC	1780	1787	Missing (in isoform 5).
FT	VARSPPLIC	1780	1787	/FTId=VSP_001368.
FT	VARIANT	314	314	V -> WTCD (IN A VARIANT).
FT	SEQUENCE	1959 AA;	208645 MW;	7FEFDFDAFF89CC31 CRC64;

Query Match 1.3%, Score 9, DB 1; Length 1959;
 Best Local Similarity 100.0%, Pred. No. 4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 466 PCLHGTCQ 474
 |||||
 Db 1228 PCLHGTCQ 1236

RESULT 4
 NTCL_MOUSE
 ID NTCL_MOUSE STANDARD; PRT; 2531 AA.
 AC Q01705; Q06007; Q61905; Q99JC2; Q9QW58; Q9ROX7;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Notch A) (mrl4) (p300).
 GN NOTCH1 OR NOTCH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RL [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Embryo;
 RX MEDLINE=93194170; PubMed=8449489;
 RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
 RA Copeland N.G., Gridley T.;
 RT "Cloning, analysis and chromosomal localization of Notch-1, a mouse
 RT homolog of Drosophila Notch.";
 RL Genomics 15:259-264(1993).
 RN [2]
 RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
 RC STRAIN=CD-1; TISSUE=Embryo;
 RX MEDLINE=93050804; PubMed=1426644;
 RA Reaume A.G., Conlon R.A., Zirngibl R., Yamaguchi T.P., Rossant J.;
 RT "Expression, analysis of a Notch homologue in the mouse embryo.";
 RL Dev. Biol. 154:377-387(1992).
 RN [3]
 RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
 RC TISSUE=Embryo;
 RX MEDLINE=93048835; PubMed=1425352;
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
 RA Greenspan R.J., McMahor A.P., Gridley T.;
 RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
 RT suggests an important role in early postimplantation mouse
 RT development.";
 RL Development 115:737-744(1992).
 RN [4]
 RP SEQUENCE OF 1161-1547 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
 RX MEDLINE=93179569; PubMed=8440332;
 RA Lardelli M., Lendahl U.;
 RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
 RT wide variety of tissues.";
 RL Exp. Cell Res. 204:364-372(1993).
 RN [5]
 RP SEQUENCE OF 1659-1673 FROM N.A.
 RX MEDLINE=99364499; PubMed=10437788;
 RA Lee J.S., Isumoto A., Yanagawa S.I.;
 RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads
 RT to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";
 RL FEBS Lett. 455:276-280(1999).
 RN [6]
 RP SEQUENCE OF 1950-2201 FROM N.A.
 RX MEDLINE=98029495; PubMed=9384671;
 RA Messerle M., Folio M., Nehls W., Eggert H., Boehm T.;
 RT "Dynamic changes in gene expression during in vitro differentiation of
 RT mouse embryonic stem cells.";
 RL Cytokines Cell. Mol. Ther. 1:1139-143(1995).
 RN [7]

RP SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
 RP MUTAGENESIS OF 1651-ARG--ARG-1654.
 RX MEDLINE=98318619; PubMed=9653148;
 RA Logeat F., Bessia C., Brou C., LeBail O., Jarriault S., Seidah N.G.,
 RA Israel A.;
 RT "The Notch1 receptor is cleaved constitutively by a furin-like
 RT convertase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
 RN [8]
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (NI-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [9]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 RN [10]
 RP INTERACTION WITH DTX1 AND DTX2.
 RX MEDLINE=21123790; PubMed=11226752;
 RA Kishi N., Tang Z., Maeda Y., Hirai A., Mo R., Ito M., Suzuki S.,
 RA Nakao K., Kinoshita T., Kadesch T., Hui C.-C., Artavanis-Tsakonas S.,
 RA Okano H., Matsuno K.;
 RT "Murine homologs of deltex define a novel gene family involved in
 RT vertebrate Notch signaling and neurogenesis.";
 RL Int. J. Dev. Neurosci. 19:21-35(2001).
 CC -!
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May play an essential role in
 CC postimplantation development, probably in some aspect of cell
 CC specification and/or differentiation. May be involved in mesoderm
 CC maturation, somitogenesis and neurogenesis. Involved in the
 CC maturation of both CD4+ and CD8+ cells in the thymus.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds. Interacts with DTX1 and DTX2.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q01705-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q01705-2; Sequence=VSP_001402, VSP_001403, VSP_001404;
 CC Note=NO experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Highly expressed in the brain, lung and
 CC thymus. Expressed at lower levels in the spleen, bone-marrow,
 CC spinal cord, eyes, mammary gland, liver, intestine, skeletal
 CC muscle, kidney and heart.
 CC -!- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 dpc By
 CC 8.5 dpc highly expressed in presomitic mesoderm, mesenchyme and
 CC endothelial cells, while much lower levels are seen in the
 CC neuroepithelium. Between 9.5-10.5 dpc expressed at high levels in
 CC the neuroepithelium. At 13.5 dpc expressed in the surface
 CC ectoderm, eye and developing whisker follicles.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a

CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -!- PTM: Phosphorylated.
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 36 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 5 ANK repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.iab-sib.ch/announce/>
 CC or send an email to license@iab-sib.ch).
 CC -----

DR EMBL; Z11886; CAA7941.1; -
 DR EMBL; L02613; AAK14898.1; -
 DR EMBL; X68278; CAA48339.1; -
 DR EMBL; AJ238029; CAB40733.1; -
 DR EMBL; X82562; CAA57909.1; -
 DR PIR; A46019; A46019.
 DR PIR; B49175; B49175.
 DR HSSP; P00740; 1EDM.
 DR MGP; MGI:97363; Notch1.
 DR GO; GO:0005887; C:integral to plasma membrane; IC.
 DR GO; GO:0005515; P:protein binding; IPI.
 DR GO; GO:0030154; P:cell differentiation; IMP.
 DR GO; GO:0007386; P:compartment specification; IMP.
 DR GO; GO:0007219; P:N signaling pathway; IC.
 DR GO; GO:0045944; P:positive regulation of transcription from P. . . ; IDA.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 7.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGF_BLOOD.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 24.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS50026; EGF_3; 36.
 DR PROSITE; PS01187; EGF_CA; 21.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
 KW Alternative splicing.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
 FT CHAIN 1711 2531 NOTCH EXTRACELLULAR TRUNCATION.
 FT CHAIN 1744 2531 NOTCH INTRACELLULAR DOMAIN.
 FT DOMAIN 19 1725 EXTRACELLULAR (POTENTIAL).

Query Match 1.3%; Score 9; DB 1; Length 2531;
 Best Local Similarity 100.0%; Pred. No. 5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 466 PCLHGTCQ 474
 |||||

Db 1029 PCLHGTCQ 1037
 RESULT 5
 NTC1 RAT
 ID NTC1_RAT STANDARD; PRT; 2531 AA.
 AC 007008;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).
 GN NOTCH1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=Schwann cell;
 RA MEDLINE=92111383; PubMed=1764995;
 RX Weinmaster G., Roberts V.J., Lemke G.;
 RT "A homolog of Drosophila Notch expressed during mammalian
 RL development."; 113:199-205 (1991).
 RL [2]
 RP REVISIONS TO 1652-1653.
 RA Weinmaster G.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP FUNCTION.
 RX MEDLINE=21094508; PubMed=1182080;
 RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
 RT Honjo T.;
 RL "Notchi and Notch3 instructively restrict bFGF-responsive multipotent
 RT neural progenitor cells to an astroglial fate.,";
 RN Neuron 29:45-55 (2001).
 RN [4]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=93202015; PubMed=1295745;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RT "Notch2: a second mammalian Notch gene.,";
 RL Development 116:931-941 (1992).
 RN [5]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21331789; PubMed=1438922;
 RA Irwin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notchi, Notch2, and Notch3 suggest multiple
 RL functional roles for the Notch-DSL signaling system during brain
 RL development.,";
 RL J. Comp. Neurol. 436:167-181 (2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC REP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). Acts instructively to control
 CC the cell fate determination of CNS multipotent progenitor cells,
 CC resulting in astroglial induction and neuron/oligodendrocyte
 CC suppression.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus (By
 CC similarity).
 CC -!- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.
 CC Expressed in postnatal central nervous system (CNS) germinal zones
 CC and, in early postnatal life, within numerous cells throughout the
 CC CNS. Found in both subventricular and ventricular germinal zones.
 CC -!- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between
 CC days 12 and 14 and decrease rapidly to much lower levels in the
 CC adult.

-!- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase in the trans-Golgi network before it reaches the plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment (NTM) and an N-terminal fragment N(BC). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane (By similarity).

-!- PTM: Phosphorylated (By similarity).

-!- SIMILARITY: Belongs to the NOTCH family.

-!- SIMILARITY: Contains 36 EGF-like domains.

-!- SIMILARITY: Contains 3 Lin/Notch repeats.

-!- SIMILARITY: Contains 5 ANK repeats.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to licens@sib-sib.ch).

EMBL; X57405; CAA40667.1; -
HSP; P00740; IEDM.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001152; Abx hydroxyl_s.
DR InterPro; IPR00742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch.
DR InterPro; IPR008800; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00066; notch; 3.
DR Pfam; PF00008; EGF; 35.
DR PIRSF; PIRSF02279; Notch; 1.
DR PRINTS; PR00010; EGFLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_CA; 25.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50086; ANK_REPEAT; 4.
DR PROSITE; PS00010; ASK_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 35.
DR PROSITE; PS01186; EGF_2; 26.
DR PROSITE; PS00026; EGF_3; 36.
DR PROSITE; PS01187; EGF_CA; 21.
KW Receptor; Transcription regulation; Activator; Differentiation;
KW Developmental protein; Receptor; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 1
FT CHAIN 19 2531
FT CHAIN 1711 2531
FT CHAIN 1744 2531
FT DOMAIN 19 1723
FT TRANSMEM 1724 1746
FT DOMAIN 1747 2531
FT DOMAIN 20 58
FT DOMAIN 59 99
FT DOMAIN 102 139
FT DOMAIN 140 176
FT DOMAIN 178 216
FT DOMAIN 218 255
FT DOMAIN 257 293
FT DOMAIN 295 333

DOMAIN	335	371	EGF-LIKE 9,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	372	410	EGF-LIKE 10,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	412	450	EGF-LIKE 11,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	452	488	EGF-LIKE 12,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	490	526	EGF-LIKE 13,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	528	564	EGF-LIKE 14,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	566	601	EGF-LIKE 15,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	603	639	EGF-LIKE 16,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	641	676	EGF-LIKE 17,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	678	714	EGF-LIKE 18,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	716	751	EGF-LIKE 19,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	753	789	EGF-LIKE 20,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	791	827	EGF-LIKE 21,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	829	867	EGF-LIKE 22,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	869	905	EGF-LIKE 23,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	907	943	EGF-LIKE 24,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	945	981	EGF-LIKE 25,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	983	1019	EGF-LIKE 26,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	1021	1057	EGF-LIKE 27,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	1059	1095	EGF-LIKE 28,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	1097	1143	EGF-LIKE 29,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	1145	1181	EGF-LIKE 30,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	1183	1219	EGF-LIKE 31,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	1221	1285	EGF-LIKE 32,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	1267	1305	EGF-LIKE 33,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	1307	1346	EGF-LIKE 34,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	1348	1384	EGF-LIKE 35,	CALCIUM-BINDING	(POTENTIAL)
DOMAIN	1387	1426	EGF-LIKE 36,	CALCIUM-BINDING	(POTENTIAL)
REPEAT	1445	1480	LIN/NOTCH 1,		
REPEAT	1481	1522	LIN/NOTCH 2,		
REPEAT	1523	1562	LIN/NOTCH 3,		
REPEAT	1917	1946	ANK 1,		
REPEAT	1950	1980	ANK 2,		
REPEAT	1984	2013	ANK 3,		
REPEAT	2017	2046	ANK 4,		
REPEAT	2047	2079	ANK 5,		
DOMAIN	2050	2079	POLY-ALA,		
DOMAIN	1891	1894	POLY-GLU,		
DOMAIN	2258	2261	POLY-PRO,		
DOMAIN	2497	2500	POLY-SER,		
SITE	1654	1655	CLEAVAGE BY (FURIN-LIKE PROTEASE) (BY SIMILARITY),		
DISULFID	24	37	BY SIMILARITY,		
DISULFID	31	46	BY SIMILARITY,		
DISULFID	48	57	BY SIMILARITY,		
DISULFID	63	74	BY SIMILARITY,		
DISULFID	68	87	BY SIMILARITY,		
DISULFID	89	98	BY SIMILARITY,		
DISULFID	106	117	BY SIMILARITY,		
DISULFID	111	127	BY SIMILARITY,		
DISULFID	129	138	BY SIMILARITY,		
DISULFID	144	155	BY SIMILARITY,		
DISULFID	149	164	BY SIMILARITY,		
DISULFID	166	175	BY SIMILARITY,		
DISULFID	182	195	BY SIMILARITY,		
DISULFID	189	204	BY SIMILARITY,		
DISULFID	206	215	BY SIMILARITY,		
DISULFID	222	233	BY SIMILARITY,		
DISULFID	227	243	BY SIMILARITY,		

Query Match 1.3%; Score 9; DB 1; Length 2531;
Best Local Similarity 100.0%; Pred.No.5;
Matches 9; Conservative 0; Mismatches 0; Gaps 0;
QY 466 PCLHGTCQ 474
Db 1029 PCLHGTCQ 1037
RESULT 6
CYSD_MYCTU STANDARD; 309 AA.
ID CYSD MYCTU
AC Q10599;

DT 16-OCT-2001 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Sulfate adenylyltransferase subunit 2 (EC 2.7.7.4) (Sulfate adenylylate transferase) (SAR) (ATP-sulfurylase small subunit).
 GN CYSD OR RV1285 OR MT1323 OR MTCY373.04 OR MB1316.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gab S., Barry C.E. III, Tekala F., Badoock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S., Hornsby T., Jagsis K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrall B.G.,
 RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickley E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";
 RL J. Bacteriol. 184:5479-5490(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H., Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S., Harris B., Atkin R., Doggett J., Maves R., Keating L., Wheeler P.R., Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 CC -1- CATALYTIC ACTIVITY: ATP + sulfate = diphosphate + adenylylsulfate.
 CC reaction occurs early in the sulfate activation pathway. This
 CC biosynthetic pathway
 CC -1- SUBUNIT: Heterodimer composed of cysD, the smaller subunit, and
 CC cysN (By similarity).
 CC -1- SIMILARITY: Belongs to the PAPS reductase family. Cysd subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL; Z73419; CAA97751.1; ALT_INIT.
 DR EMBL; AE007007; AAK45584.1; -
 DR EMBL; BX248338; CAD94177.1; ALT_INIT.
 DR TrEMBL; MT1323; -
 DR Tuberculist; Rv1285; -
 DR HAMAP; MF_00064; -; 1.
 DR InterPro; IPR002500; PAPS_reduct.
 DR Pfam; PF01507; PAPS_reduct; 1.
 DR Cysteine biosynthesis; Transferase; Nucleotidyltransferase;

KW Complete proteome.
 SQ SEQUENCE 309 AA; 34888 MW; 75D965D35F1EC284 CRC64;
 Query Match 1.1%; Score 8; DB 1; Length 309;
 Best Local Similarity 100.0%; Fred. No. 7.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 279 VLLFSGGK 286
 Db 41 VLLFSGGK 48
 RESULT 7
 AMPA_EACSU STANDARD; PRT; 500 AA.
 ID AMPA_EACSU
 AC Q32106;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase) (LAP) (Leucyl aminopeptidase).
 GN PEPA OR BSU32050.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., S., Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borrillet S., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniels R.A., Denzot F., Davine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G., Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle B., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Soldo B., Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B., Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terptrtra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vaasarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzengger T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zmstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis";
 RL Nature 390:249-256(1997).
 CC -1- FUNCTION: Presumably involved in the processing and regular turnover of intracellular proteins. Catalyzes the removal of unsubstituted N-terminal amino acids from various peptides (By similarity)
 CC -1- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-Xbb-, in which xaa is preferably Leu, but may be other amino acids including Pro although not Arg or Lys, and Xbb may be Pro.
 CC -1- COFACTOR: Binds 2 manganese ions per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to peptidase family M17.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC -----
 CC EMBL; Z99120; CAB15195.1; -
 CC PIR; F70012; F70012.
 CC HSSP; P00727; 1LAM.
 CC MEROPS; M17.UFW; -
 CC Subtilast; BG13970; pepA.
 CC HAMAP; MF_00181; -; 1.
 CC InterPro; IPR000819; Peptidase M17 C.
 CC InterPro; IPR008283; Peptidase M17_N.
 CC Pfam; PF00883; Peptidase M17; 1.
 CC Pfam; PF02789; Peptidase M17_N; 1.
 CC PRINTS; PRO0481; LAMNOPTDASE.
 CC PROSITE; PS00631; CYTOSOL_AP; 1.
 KW Hydroxylase; Aminopeptidase; Manganese; Complete proteome.
 FT METAL 261
 FT METAL 266
 FT METAL 284
 FT METAL 284
 FT METAL 343
 FT METAL 345
 FT ACT_SITE 273
 FT ACT_SITE 347
 FT ACT_SITE 347
 SQ SEQUENCE 500 AA; 53657 MW; 3E82968F656559 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 500;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 653 DVATLTGG 660
 DB 367 DVATLTGG 374

RESULT 8
 PGCN_RAT STANDARD; PRT; 1257 AA.
 AC P55067;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3)
 DE (245 kDa early postnatal core glycoprotein) [Contains: 150 kDa adult
 DE core glycoprotein].
 GN CSPG3 OR NCAN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116,
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=92406907; PubMed=1326557;
 RA Rauch U., Karthikeyan L., Maurel P., Margolis R.U., Margolis R.K.;
 RT "Cloning and primary structure of neurocan, a developmentally
 RT regulated aggregating chondroitin sulfate proteoglycan of brain.";
 RL J Biol. Chem. 267:19536-19547(1992).
 RN [2]

RP CHARACTERIZATION.
 RX MEDLINE=94230574; PubMed=7513709;
 RA Friedlander D.R., Milev P., Karthikeyan L., Margolis R.K.,
 RA Margolis R.U., Grumet M.;
 RT "The neuronal chondroitin sulfate proteoglycan neurocan binds to the
 RT neural cell adhesion molecules Ng-CAM/LL/NILE and N-CAM, and inhibits
 RT neuronal adhesion and neurite outgrowth.";
 RL J. Cell Biol. 125:669-680(1994).
 CC -1- FUNCTION: May modulate neuronal adhesion and neurite growth during
 CC development by binding to neural cell adhesion molecules (NG-CAM
 CC and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
 CC acid.

CC -1- TISSUE SPECIFICITY: Early postnatal and adult brain; not expressed
 CC in kidney, lung, liver and muscle.
 CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
 CC O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
 CC -1- PTM: TWO ISOFORMS WERE FOUND THAT PROBABLY ARISE BY PROTEOLYTIC
 CC DEGRADATION. THE LARGE ISOFORM IS PREDOMINANT IN EARLY POSTNATAL
 CC BRAIN, THE SMALL ISOFORM IS FOUND IN ADULT BRAIN.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -1- SIMILARITY: Contains 2 EGF-like domains.
 CC -1- SIMILARITY: Contains 2 link domains.
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -1- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC -----
 CC EMBL; M97161; AAC37679.1; -
 CC PIR; S28764; S28764.
 CC HSSP; P00740; 1EDW.
 CC InterPro; IPR002353; AntifreezeII.
 CC InterPro; IPR000152; Asx hydroxyl_s.
 CC InterPro; IPR000742; EGF_2
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR006209; EGF-Like.
 CC InterPro; IPR007110; IG-Like.
 CC InterPro; IPR003599; IG.
 CC InterPro; IPR001304; Lectin C.
 CC InterPro; IPR000538; Link.
 CC InterPro; IPR000436; Sushi_SCR_CCP.
 CC Pfam; PF00008; EGF; 2.
 CC Pfam; PF00047; ig; 1.
 CC Pfam; PF00059; lectin_C; 1.
 CC Pfam; PF00084; sushi; 1.
 CC Pfam; PF00193; xlink; 2.
 CC PRINTS; PR00356; ANTIFREEZEII.
 CC PRINTS; PR01265; LINKMODULE.
 CC ProDom; PD000918; Link; 2.
 CC SMART; SM00032; CCP; 1.
 CC SMART; SM00034; CUECT; 1.
 CC SMART; SM00179; EGF_CA; 1.
 CC SMART; SM00409; IG; 1.
 CC SMART; SM00445; LINK; 2.
 CC PROSITE; PS00010; ASX HYDROXYL; 1.
 CC PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
 CC PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
 CC PROSITE; PS00022; EGF_1; 3.
 CC PROSITE; PS01186; EGF_2; 1.
 CC PROSITE; PS50026; EGF_3; 2.
 CC PROSITE; PS01187; EGF_CA; 1.
 CC PROSITE; PS50835; IG_LIKE; 1.
 CC PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
 KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
 FT CHAIN 1 22
 FT CHAIN 23 1257
 FT CHAIN 639 1257
 FT DOMAIN 37 157
 FT DOMAIN 158 253
 FT DOMAIN 259 355
 FT DOMAIN 949 985
 FT DOMAIN 987 1023
 FT DOMAIN 1025 1154
 FT DOMAIN 1155 1213
 FT DOMAIN 58 139
 FT DISULFID 181 252
 FT DISULFID 205 226
 FT DISULFID 279 354

NEUROCAN CORE PROTEIN.
 150 kDa ADULT CORE GLYCOPROTEIN.
 IG-LIKE V-TYPE.
 LINK 1.
 LINK 2.
 EGF-LIKE 1.
 EGF-LIKE 2.
 C-TYPE LECTIN.
 SUSHI.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.

FT DISULFID 302 BY SIMILARITY.
 FT DISULFID 303 BY SIMILARITY.
 FT DISULFID 304 BY SIMILARITY.
 FT DISULFID 978 BY SIMILARITY.
 FT DISULFID 979 BY SIMILARITY.
 FT DISULFID 1029 BY SIMILARITY.
 FT DISULFID 1057 BY SIMILARITY.
 FT DISULFID 1125 BY SIMILARITY.
 FT DISULFID 1156 BY SIMILARITY.
 FT DISULFID 1185 BY SIMILARITY.
 FT CARBOHYD 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 944 O-LINKED (XYL. . .) (CHONDROITIN SULFATE).
 FT CARBOHYD 967 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1164 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1257 AA; 135544 MW; 992B33DC9FAL9EBB CRC64;

Query Match 1.1%; Score 8; DB 1; Length 1257;
 Best Local Similarity 100.0%; Pred.No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGGTC 473
 DB 957 PCLHGGTC 964

RESULT 9
 PGCN_MOUSE STANDARD; PRT; 1268 AA.
 AC P55056;
 ID PGCN_MOUSE
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
 GN CSPG3 OR NCAN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCSI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C; TISSUE=Brain;
 RX MEDLINE=96039250; PubMed=7490074;
 RA Rauch U., Grimpe B., Kalbe G., Arnold-Ammer I., Beier D.,
 RA Passler R.;
 RT Structure and chromosomal localization of the mouse neurocan gene.;
 RL Genomics 28:405-410(1995).
 CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during development by binding to neural cell adhesion molecules (NG-CAM and N-CAM). Chondroitin sulfate proteoglycan, binds to hyaluronic acid
 CC -!- TISSUE SPECIFICITY: Brain.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X84727; CAA59216.1;
 CC PIR; S52781; S52781.
 CC HSSP; P00740; IEDM.
 CC MGD; MGI:104694; Cspg3.
 CC InterPro; IPR002353; AntifreezeZell.
 CC InterPro; IPR000152; Asx_hydroxyl_S.

DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR007110; Ig_Like.
 DR InterPro; IPR003599; Ig_Like.
 DR InterPro; IPR001394; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF_2.
 DR Pfam; PF00047; Ig_1.
 DR Pfam; PF00059; lectin_c; 1.
 DR Pfam; PF00084; sushi_1.
 DR Pfam; PF00193; Xlink; 2.
 DR PRINTS; PRO0356; ANTIFREEZEL1.
 DR PRINTS; PRO1265; LINKMODULE.
 DR PRODOM; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00409; Ig_1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
 DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1_3.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS00026; EGF_2; 2.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS00835; Ig_Like; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
 KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
 FT SIGNAL 1 22 POTENTIAL
 FT CHAIN 23 1268 NEUROCAN CORE PROTEIN.
 FT DOMAIN 37 157 IG-LIKE V-TYPE.
 FT DOMAIN 158 253 LINK 1.
 FT DOMAIN 259 355 LINK 2.
 FT DOMAIN 960 996 EGF-LIKE 1.
 FT DOMAIN 998 1034 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1036 1165 C-TYPE LECTIN.
 FT DOMAIN 1166 1224 SUSHI.
 FT DISULFID 58 139 BY SIMILARITY.
 FT DISULFID 181 252 BY SIMILARITY.
 FT DISULFID 205 226 BY SIMILARITY.
 FT DISULFID 279 354 BY SIMILARITY.
 FT DISULFID 303 324 BY SIMILARITY.
 FT DISULFID 964 975 BY SIMILARITY.
 FT DISULFID 969 984 BY SIMILARITY.
 FT DISULFID 986 995 BY SIMILARITY.
 FT DISULFID 1040 1051 BY SIMILARITY.
 FT DISULFID 1068 1160 BY SIMILARITY.
 FT DISULFID 1136 1152 BY SIMILARITY.
 FT DISULFID 1167 1210 BY SIMILARITY.
 FT DISULFID 1196 1223 BY SIMILARITY.
 FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1268 AA; 137200 MW; 3014E8E202A2FAEC CRC64;

Query Match 1.1%; Score 8; DB 1; Length 1268;
 Best Local Similarity 100.0%; Pred.No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGGTC 473
 DB 968 PCLHGGTC 975

RESULT 10
 ID PGCN_HUMAN STANDARD; PRT; 1321 AA.

AC O14594; Q9UPK6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
 GN CSPG3 OR NCAN OR NEUR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99013874; PubMed=9795216;
 RX Prange C.K., Pennacchio L.A., Lieuallen K., Fan W., Leunon G.G.;
 RT "Characterization of the human neurocan gene, CSPG3.";
 RL Gene 221:199-205(1998).
 RN [2]
 RP SEQUENCE OF 1-990 AND 1007-1321 FROM N.A., AND VARIANT ALA-1254.
 RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
 RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
 RA Phan H., Velasco N., Do L., Regala M., Terry A., Gaines J.,
 RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
 RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
 RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
 RA Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
 RA Kobayashi A., Olsen A.S., Carrano A.V.;
 RT "Sequence analysis of an ~1 Mb region containing the MEF2B gene in
 RT 19p12.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBSJ databases.
 CC !- FUNCTION: May modulate neuronal adhesion and neurite growth during
 CC development by binding to neural cell adhesion molecules (NG-CAM
 CC and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronin
 CC acid.
 CC !- TISSUE SPECIFICITY: Brain.
 CC !- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC !- SIMILARITY: Contains 2 EGF-like domains.
 CC !- SIMILARITY: Contains 2 link domains.
 CC !- SIMILARITY: Contains 1 C-type lectin family domain.
 CC !- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC !- SIMILARITY: Belongs to the aggregan/versican proteoglycan family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announcement/>
 CC or send an email to license@ebi-sib.ch).
 CC -----
 DR EMBL; AF036547; AAC80576.1; -
 DR EMBL; AC003110; AAB86655.1; -
 DR EMBL; AC005254; AAC25581.1; -
 DR HSSP; P00740; 1EDM.
 DR Genew; HGNC:2465; CSPG3.
 DR MIM; 600826; -
 DR InterPro; IPR00152; Asx hydroxyl_s.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR007110; IG-Ilike.
 DR InterPro; IPR003599; IG.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR005338; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00047; ig; 1.
 DR Pfam; PF00059; lectin_c; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 1.
 DR PRINTS; PR01265; LINKMODULE.
 DR ProDom; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.

DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX HYDROXYL; 1.
 DR PROSITE; PS00615; C TYPE LECTIN_1; 1.
 DR PROSITE; PS00041; C TYPE LECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1_3.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS00026; EGF_3; 2.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS05083; IG LIKE; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
 KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal; Polymorphism.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 1321 NEUROCAN CORE PROTEIN.
 FT DOMAIN 38 153 IG-LIKE V-TYPE.
 FT DOMAIN 159 254 LINK 1.
 FT DOMAIN 260 356 LINK 2.
 FT DOMAIN 1008 1044 EGF-LIKE 1.
 FT DOMAIN 1045 1082 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1084 1213 C-TYPE LECTIN.
 FT DOMAIN 1214 1272 SUSHI.
 FT DISULFID 59 140 BY SIMILARITY.
 FT DISULFID 182 253 BY SIMILARITY.
 FT DISULFID 206 227 BY SIMILARITY.
 FT DISULFID 280 355 BY SIMILARITY.
 FT DISULFID 304 325 BY SIMILARITY.
 FT DISULFID 1012 1023 BY SIMILARITY.
 FT DISULFID 1017 1032 BY SIMILARITY.
 FT DISULFID 1034 1043 BY SIMILARITY.
 FT DISULFID 1088 1099 BY SIMILARITY.
 FT DISULFID 1116 1208 BY SIMILARITY.
 FT DISULFID 1184 1200 BY SIMILARITY.
 FT DISULFID 1215 1258 BY SIMILARITY.
 FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1026 1026 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1223 1223 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 1254 1254 V -> A (in dbSNP:1064389).
 FT CONFLICT 1234 1234 N -> Y (IN REF. 2).
 FT CONFLICT 1282 1282 G -> R (IN REF. 2).
 FT SEQUENCE 1321 AA; 142972 MW; 2EF47F823DB980B8 CRC64;
 SQ
 Query Match 1.1%; Score 8; DB 1; Length 1321;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 466 PCLHGTC 473
 Db 1016 PCLHGTC 1023
 RESULT 11
 AGRI_DISOM
 ID AGRI_DISOM STANDARD; PRT; 1328 AA.
 AC Q90404;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Agrin (Fragment).
 GN AGRN
 OS Discoptys omata (Electric ray).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squalia; Hyposquales; Pristiogalea; Batoida;
 OC Torpediniformes; Narcinoidei; Narcinidae; Discoptys.
 OX NCBI_TaxID=7785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smith M.A., Magill-Solc C., Rupp F., Yao Y.-M.M., Schilling J.W.,
 RA Snow P., McMahan U.J.;

RT "Isolation and characterization of a cDNA that encodes an agrin homolog in the marine ray."
 RL Mol. Cell. Neurosci. 3:406-417(1992).
 CC -!- FUNCTION: Component of the basal lamina that causes the aggregation of acetylcholine receptors and acetylcholine-esterase on the surface of muscle fibers of the neuromuscular junction (By similarity).
 CC -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular junction (By similarity).
 CC -!- SIMILARITY: Contains at least 2 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains at least 3 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 SEA domain.
 CC -!- SIMILARITY: Contains 3 laminin G-like domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L01423; AAA49224.1; -
 DR PIR; T43060; T43060.
 DR HSSP; P00740; LEDM.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR008985; ConA_like_1ec_s1.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR008209; EGF_2.
 DR InterPro; IPR003645; FOLN.
 DR InterPro; IPR002350; kazal.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF00008; EGF_4.
 DR Pfam; PF00050; kazal; 2.
 DR Pfam; PF00053; laminin_EGF; 2.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF01390; SEA; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR SMART; SM00180; EGF_Lam; 2.
 DR SMART; SMO0274; FOLN; 2.
 DR SMART; SMO0280; KAZAL; 2.
 DR SMART; SMO0282; LamG; 3.
 DR SMART; SMO0200; SEA; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; 5.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50026; EGF_3; 4.
 DR PROSITE; PS50025; Lam_G_DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 DR PROSITE; PS50024; SEA; 1.
 KW Glycoprotein; EGF-like domain; Repeat; Laminin EGF-like domain.
 FT NON_TER 1 1
 FT DOMAIN 79 132 LAMININ EGF-LIKE 1.
 FT DOMAIN 133 179 LAMININ EGF-LIKE 2.
 FT DOMAIN 411 533 SEA.
 FT DOMAIN 608 644 EGF-LIKE 1.
 FT DOMAIN 649 825 LAMININ G-LIKE 1.
 FT DOMAIN 865 902 EGF-LIKE 2.
 FT DOMAIN 914 1096 LAMININ G-LIKE 2.
 FT DOMAIN 1097 1135 EGF-LIKE 3.
 FT DOMAIN 1146 1324 LAMININ G-LIKE 3.
 FT DISULFID 79 91 BY SIMILARITY.
 FT DISULFID 81 98 BY SIMILARITY.
 FT DISULFID 100 109 BY SIMILARITY.
 FT DISULFID 112 130 BY SIMILARITY.
 FT DISULFID 133 145 BY SIMILARITY.
 FT DISULFID 135 152 BY SIMILARITY.
 FT DISULFID 154 163 BY SIMILARITY.
 FT DISULFID 166 177 BY SIMILARITY.
 FT DISULFID 612 623 BY SIMILARITY.
 FT DISULFID 617 632 BY SIMILARITY.

FT DISULFID 634 643 BY SIMILARITY.
 FT DISULFID 869 880 BY SIMILARITY.
 FT DISULFID 874 890 BY SIMILARITY.
 FT DISULFID 892 901 BY SIMILARITY.
 FT DISULFID 1101 1114 BY SIMILARITY.
 FT DISULFID 1108 1123 BY SIMILARITY.
 FT DISULFID 1125 1134 BY SIMILARITY.
 FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 369 369 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1328 AA; 144018 MW; 79D81CIAF2A71C18 CRC64;
 Query Match 1..1; Score 8; DB 1; Length 1328;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 466 PCLHGTC 473
 DB 616 PCLHGTC 623
 RESULT 12
 AGRI_CHKC
 ID AGRI_CHKC STANDARD; PRT; 1955 AA.
 AC P31696;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DR Agrin precursor.
 GN AGRN.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=92232297; PubMed=1314620;
 RA Tsim K.W.K., Ruegg M.A., Escher G., Kroeger S., McMahon U.J.;
 RT "cDNA that encodes active agrin."
 RL Neuron 8:677-689(1992).
 RP [2]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=92232298; PubMed=1314621;
 RA Ruegg M.A., Tsim K.W.K., Horton S.E., Escher G.,
 RA Gensch E.M., McMahon U.J.;
 RT "The agrin gene codes for a family of basal lamina proteins that differ in function and distribution."
 RL Neuron 8:691-699(1992).
 CC -!- FUNCTION: Component of the basal lamina that causes the aggregation of acetylcholine receptors and acetylcholine-esterase on the surface of muscle fibers of the neuromuscular junction.
 CC -!- SUBUNIT: Binds to laminin.
 CC -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular junction.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Additional isoforms seem to exist. Isoforms differ in their acetylcholine receptor clustering activity;
 CC Name=1;
 CC IsoId=P31696-1; Sequence=Displayed;
 CC Name=2; Synonyms=Agrin-related protein 1;
 CC IsoId=P31696-2; Sequence=VSP_001370;
 CC Name=3; Synonyms=Agrin-related protein 2;
 CC IsoId=P31696-3; Sequence=VSP_001369; VSP_001370;
 CC -!- FTM: Contains heparan sulfate chains as well as N-linked and O-linked oligosaccharides (By similarity).
 CC -!- SIMILARITY: Contains 9 Kazal-like domains.
 CC -!- SIMILARITY: Contains 2 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 4 EGF-like domains.

FT	DISULFID	789	BY SIMILARITY.
FT	DISULFID	806	BY SIMILARITY.
FT	DISULFID	811	BY SIMILARITY.
FT	DISULFID	826	BY SIMILARITY.
FT	DISULFID	837	BY SIMILARITY.
FT	DISULFID	844	BY SIMILARITY.
FT	DISULFID	855	BY SIMILARITY.
FT	DISULFID	890	BY SIMILARITY.
FT	DISULFID	901	BY SIMILARITY.
FT	DISULFID	919	BY SIMILARITY.
FT	DISULFID	928	BY SIMILARITY.
FT	DISULFID	930	BY SIMILARITY.
FT	DISULFID	939	BY SIMILARITY.
FT	DISULFID	946	BY SIMILARITY.
FT	DISULFID	952	BY SIMILARITY.
FT	DISULFID	957	BY SIMILARITY.
FT	DISULFID	968	BY SIMILARITY.
FT	DISULFID	984	BY SIMILARITY.
FT	DISULFID	989	BY SIMILARITY.
FT	DISULFID	1009	BY SIMILARITY.
FT	DISULFID	1011	BY SIMILARITY.
FT	DISULFID	1020	BY SIMILARITY.
FT	DISULFID	1211	BY SIMILARITY.
FT	DISULFID	1212	BY SIMILARITY.
FT	DISULFID	1231	BY SIMILARITY.
FT	DISULFID	1233	BY SIMILARITY.
FT	DISULFID	1242	BY SIMILARITY.
FT	DISULFID	1485	BY SIMILARITY.
FT	DISULFID	1496	BY SIMILARITY.
FT	DISULFID	1490	BY SIMILARITY.
FT	DISULFID	1505	BY SIMILARITY.
FT	DISULFID	1516	BY SIMILARITY.
FT	DISULFID	1767	BY SIMILARITY.
FT	DISULFID	1774	BY SIMILARITY.
FT	DISULFID	1783	BY SIMILARITY.
FT	DISULFID	1788	BY SIMILARITY.
FT	DISULFID	1794	BY SIMILARITY.
FT	DISULFID	1801	BY SIMILARITY.
FT	DISULFID	1812	BY SIMILARITY.
FT	DISULFID	1806	BY SIMILARITY.
FT	DISULFID	1821	BY SIMILARITY.
FT	DISULFID	1823	BY SIMILARITY.
FT	DISULFID	1839	BY SIMILARITY.
FT	DISULFID	1850	BY SIMILARITY.
FT	DISULFID	1844	BY SIMILARITY.
FT	DISULFID	1859	BY SIMILARITY.
FT	DISULFID	1861	BY SIMILARITY.
FT	DISULFID	1870	BY SIMILARITY.
FT	DISULFID	1878	BY SIMILARITY.
FT	DISULFID	1883	BY SIMILARITY.
FT	DISULFID	1893	BY SIMILARITY.
FT	DISULFID	1905	BY SIMILARITY.
FT	DISULFID	1914	BY SIMILARITY.
FT	DISULFID	1919	BY SIMILARITY.
FT	DISULFID	1924	BY SIMILARITY.
FT	DISULFID	1939	BY SIMILARITY.
FT	DISULFID	1941	BY SIMILARITY.
FT	DISULFID	1950	BY SIMILARITY.
FT	DISULFID	1957	BY SIMILARITY.
FT	DISULFID	1968	BY SIMILARITY.
FT	DISULFID	1962	BY SIMILARITY.

Query Match	1.1%	Score 8;	DB 1;
Best Local Similarity	100.0%	Pred. No. 44;	
Matches	8;	Conservative	0;
		Mismatches	0;
		Indels	0;
		Gaps	0;

OY	166	PCQNGQC	173
Db	1961	PCQNGQC	1668

RESULT 14
NTC3 HUMAN
ID NTC3 HUMAN STANDARD; PRT; 2321 AA.
AC Q9UM47, Q9UEB3, Q9UEP3, Q9Y6L6;
DI 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN NOTCH3.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97032728; PubMed=8878478;
RA Joutel A., Corpechot C., Ducros A., Vahedi K., Chabriat H., Mouton P.,
RA Alamowitch S., Domenga V., Cecillon M., Marechal E., Maciazek J.,
RA Vaysiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissenbach J.,
RA Bach J.-F., Bousser M.-G., Tournier-Lasserre E.;
RT Notch3 mutations in CADASIL, a hereditary adult-onset condition
RT causing stroke and dementia.";
RL Nature 383:707-710 (1996).
RN [2]

SEQUENCE FROM N.A.
 Gunel M., Artavanis-Tsakonas S.;
 Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
 [3]
 SEQUENCE FROM N.A.
 Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
 Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
 Phan H., Velasco N., Ganes J., Danganan L., Poundstone P.,
 Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
 Frankheim M., Amico-Keller G., Cosfield J., Duarte S., Lucas S.,
 Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
 Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
 Carrano A.V.;
 "Sequence analysis of an 1.5 Mb olfactory receptor (OLFRL) cluster in
 19p13.1";
 Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
 [4]
 VARIANTS CADASIL TYR-49; CYS-71; CYS-90; CYS-110; CYS-133; CYS-141;
 ARG-146; CYS-153; CYS-169; CYS-172; CYS-182; ARG-185; SER-212;
 GLY-222; TYR-224; CYS-258; TYR-542; CYS-558; CYS-578; CYS-728;
 CYS-985; CYS-1006; CYS-1031; CYS-1231 AND ARG-1261, AND VARIANTS
 ARG-170; LEU-496; GLN-1133; MET-1183 AND ALA-2223.
 MEDLINE=98049753; PubMed=9388399;
 Joutel A., Vahedi K., Corpechot C., Troesch A., Chabriat H.,
 Vaysiere C., Cruaud C., Maciazek J., Weissenbach J., Bousser M.-G.,
 Bach J.-F., Tournier-Lasserre E.;

"Strong clustering and stereotyped nature of Notch3 mutations in
 CADASIL patients";
 RT Lancet 350:1511-1515 (1997).
 RL [5]
 VARIANT CADASIL 114-GLY--PRO-120 DEL.
 RP MEDLINE=20264473; PubMed=10802807;
 RX Joutel A., Chabriat H., Vahedi K., Domenga V., Vaysiere C.,
 Ruchoux M.M., Lucas C., Leys D., Bousser M.-G., Tournier-Lasserre E.;

"Splice site mutation causing a seven amino acid Notch3 in-frame
 deletion in CADASIL";
 RT Neurology 54:1874-1875 (2000).
 RL [6]
 IDENTIFICATION OF LIGANDS.
 RP MEDLINE=99180765; PubMed=10079256;
 RX Grays G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;

"Human ligands of the Notch receptor.";
 RT Am. J. Pathol. 154:785-794 (1999).
 RL [7]
 FUNCTION: Functions as a receptor for membrane-bound ligands
 Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 Upon ligand activation through the released notch intracellular
 domain (NICD) it forms a transcriptional activator complex with
 RBP-J kappa and activates genes of the enhancer of split locus.
 Affects the implementation of differentiation, proliferation and
 apoptotic programs (By similarity).
 CC [8]
 SUBUNIT: Heterodimer of a C-terminal fragment N(TM), and a N-
 terminal fragment N(EC) which are probably linked by disulfide
 bonds (By similarity).
 CC [9]
 SUBCELLULAR LOCATION: Type I membrane protein. Following
 proteolytical processing NICD is translocated to the nucleus.
 CC [10]
 TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
 tissues.
 CC [11]
 PTM: Synthesized in the endoplasmic reticulum as an inactive form
 which is proteolytically cleaved by a furin-like convertase in the
 trans-Golgi network before it reaches the plasma membrane to yield
 an active, ligand-accessible form. Cleavage results in a C-
 terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 ligand binding, it is cleaved by TNF-alpha converting enzyme
 (TACE) to yield a membrane-associated intermediate fragment called
 notch extracellular truncation (NEXT). This fragment is then
 cleaved by presenilin dependent gamma-secretase to release a
 notch-derived peptide containing the intracellular domain (NICD)
 from the membrane (By similarity).
 CC [12]
 PTM: Phosphorylated (By similarity).
 CC [13]
 DISBASE: Defects in NOTCH3 are associated with cerebral autosomal
 dominant arteriopathy with subcortical infarcts and
 leukoencephalopathy (CADASIL) [MIM:125310]. CADASIL causes a type

CC of stroke and dementia of which key features include recurrent
 CC subcortical ischemic events and vascular dementia.
 CC -1- SIMILARITY: Belongs to the NOTCH family.
 CC -1- SIMILARITY: Contains 34 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -1- SIMILARITY: Contains 5 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U97669; AAB91371.1; -
 DR EMBL; AF058900; AAC14346.1; -
 DR EMBL; AF058881; AAC14346.1; JOINED.
 DR EMBL; AF058882; AAC14346.1; JOINED.
 DR EMBL; AF058883; AAC14346.1; JOINED.
 DR EMBL; AF058884; AAC14346.1; JOINED.
 DR EMBL; AF058885; AAC14346.1; JOINED.
 DR EMBL; AF058886; AAC14346.1; JOINED.
 DR EMBL; AF058887; AAC14346.1; JOINED.
 DR EMBL; AF058888; AAC14346.1; JOINED.
 DR EMBL; AF058889; AAC14346.1; JOINED.
 DR EMBL; AF058890; AAC14346.1; JOINED.
 DR EMBL; AF058891; AAC14346.1; JOINED.
 DR EMBL; AF058892; AAC14346.1; JOINED.
 DR EMBL; AF058893; AAC14346.1; JOINED.
 DR EMBL; AF058894; AAC14346.1; JOINED.
 DR EMBL; AF058895; AAC14346.1; JOINED.
 DR EMBL; AF058896; AAC14346.1; JOINED.
 DR EMBL; AF058897; AAC14346.1; JOINED.
 DR EMBL; AF058898; AAC14346.1; JOINED.
 DR EMBL; AF058899; AAC14346.1; JOINED.
 DR EMBL; AC004257; AAC04897.1; -
 DR EMBL; AC004663; AAC15789.1; ALT_INIT.
 DR PIR; S78549; S78549.
 DR HSP; P00740; IEDM.
 DR Gnew; HGNC:7893; NOTCH3.
 DR MIM; 600276; -
 DR MIM; 125310; -
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_1.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00009; EGF; 34.
 DR Pfam; PF00086; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR00011; EGFBLAMIN.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_Ca; 19.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PSS0297; ANK_REPEAT; 1.
 DR PROSITE; PSS00088; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 18.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 25.
 DR PROSITE; PSS0026; EGF_3; 34.
 DR PROSITE; PS01187; EGF_Ca; 16.
 DR Receptor; Transcription regulation; Activator; Differentiation;
 DR KW Kewceptor; Transcription regulation; Repeat; ANK repeat; EGF-like domain;
 DR KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 DR Disease mutation.

FT SIGNAL 1 39
 FT CHAIN 40 2321
 FT CHAIN 1629 2321
 FT CHAIN 1662 2321
 FT DOMAIN 40 1643
 FT TRANSMEM 1644 1664
 FT DOMAIN 1665 2321
 FT DOMAIN 40 77
 FT DOMAIN 78 118
 FT DOMAIN 119 156
 FT DOMAIN 158 195
 FT DOMAIN 197 234
 FT DOMAIN 236 272
 FT DOMAIN 274 312
 FT DOMAIN 314 350
 FT DOMAIN 351 389
 FT DOMAIN 391 429
 FT DOMAIN 431 467
 FT DOMAIN 469 505
 FT DOMAIN 507 543
 FT DOMAIN 545 580
 FT DOMAIN 582 618
 FT DOMAIN 620 655
 FT DOMAIN 657 693
 FT DOMAIN 695 730
 FT DOMAIN 734 770
 FT DOMAIN 771 808
 FT DOMAIN 810 847
 FT DOMAIN 849 885
 FT DOMAIN 887 922

POTENTIAL.
 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
 NOTCH EXTRACELLULAR TRUNCATION (BY SIMILARITY).
 NOTCH INTRACELLULAR DOMAIN (BY SIMILARITY).
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 EGF-LIKE 1.
 EGF-LIKE 2.
 EGF-LIKE 3.
 EGF-LIKE 4.
 EGF-LIKE 5.
 EGF-LIKE 6.
 EGF-LIKE 7.
 EGF-LIKE 8.
 EGF-LIKE 9.
 EGF-LIKE 10.
 EGF-LIKE 11.
 EGF-LIKE 12.
 EGF-LIKE 13.
 EGF-LIKE 14.
 EGF-LIKE 15.
 EGF-LIKE 16.
 EGF-LIKE 17.
 EGF-LIKE 18.
 EGF-LIKE 19.
 EGF-LIKE 20.
 EGF-LIKE 21.
 EGF-LIKE 22.
 EGF-LIKE 23.

Query Match 1.1%; Score 8; DB 1; Length 2321;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 483 GFSGPRCQ 490
 |||||
 Db 1237 GFSGPRCQ 1244

RESULT 15
 NTC2_MOUSE STANDARD; PRT; 2470 AA.
 AC O35516; Q06008; Q60941;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch B).
 GN NOTCH2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Thymus;
 RA Hamada Y., Higuchi M., Tsujimoto Y.;
 RT "Complete amino acid sequence and multiform transcripts encoded by a single copy of mouse Notch2 gene";
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 316-1518 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
 RX MARDINE=93178563; PubMed=8440332;
 RA Lardelli M., Lendahl U.;
 RT "Motch A and Motch B-two mouse Notch homologues coexpressed in a wide variety of tissues.";
 RL Exp. Cell Res. 204:364-372(1993).
 RN [3]
 RP SEQUENCE OF 1765-2153 FROM N.A.

RX MEDLINE=97075110; PubMed=8917536;
 RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
 RA Martin D.I.;
 RT "Inhibition of granulocytic differentiation by mNotch1."
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
 RN [4]
 RP FUNCTION.
 RA MEDLINE=99396706; PubMed=10393120;
 RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
 RA Tsujimoto Y.;
 RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
 RT embryonic lethality."
 RL Development 126:3415-3424(1999).
 RN [5]
 RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
 RA MEDLINE=95333893; PubMed=7609614;
 RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
 RT "Differential expression of Notch1 and Notch2 in developing and adult
 RT mouse brain."
 RL Brain Res. Mol. Brain Res. 29:263-272(1995).
 RN [6]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RA MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis."
 RL J. Biol. Chem. 276:40289-40273(2001).
 RN [7]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RA MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May play an essential role in
 CC postimplantation development, probably in some aspect of cell
 CC specification and/or differentiation.
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=O35516-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=O35516-2; Sequence=Vsp 001405;
 CC Note=No experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
 CC neuroepithelia, somites, optic vesicles and branchial arches, but
 CC not heart.
 CC -!- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
 CC the postnatal ependymal cells, and the choroid plexus throughout
 CC embryonic and postnatal development.
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -!- PTM: Phosphorylated.

CC CC
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 35 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 6 ANK repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D32210; BAA22094.1; -
 CC EMBL; X68279; CAA48340.1; -
 CC EMBL; U31881; AAC52924.1; -
 CC FRR; A49175; A49175.
 CC HSP; P16109; IFSB.
 CC MGD; MGI:97364; Notch2.
 CC GO; GO:0005887; C:integral to plasma membrane; IC.
 CC GO; GO:0005515; F:protein binding; IPI.
 CC GO; GO:002011; P:morphogenesis of an epithelial sheet; IMP.
 CC GO; GO:0007219; P:N signaling pathway; IC.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl_s.
 CC InterPro; IPR000742; EGF 2.
 CC InterPro; IPR001881; EGF Ca.
 CC InterPro; IPR001438; EGF II.
 CC InterPro; IPR006209; EGF like.
 CC InterPro; IPR002049; Laminin_EGF.
 CC InterPro; IPR008297; Notch.
 CC InterPro; IPR000690; Notch_dom.
 CC Pfam; PF00023; ANK; 6.
 CC Pfam; PF00008; EGF; 34.
 CC Pfam; PF00066; notch; 2.
 CC PIRSF; PIRSF002279; Notch; 1.
 CC PRINTS; PR00010; EGFLOOD.
 CC PRINTS; PR00011; EGFAMININ.
 CC PRINTS; PRO1452; NOTCH.
 CC SMART; SM00248; ANK; 6.
 CC SMART; SM00179; EGF_CA; 23.
 CC SMART; SM00004; NL; 3.
 CC PROSITE; PS50297; ANK REP REGION; 1.
 CC PROSITE; PS50088; ANK_REPEAT; 4.
 CC PROSITE; PS00010; ASX_HYDROXYL; 22.
 CC PROSITE; PS00022; EGF_1; 33.
 CC PROSITE; PS01186; EGF_2; 27.
 CC PROSITE; PS50026; EGF_3; 35.
 CC PROSITE; PS01187; EGF_CA; 22.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation;
 CC Alternative splicing.
 CC SIGNAL 1 25
 CC CHAIN 26 2470 POTENTIAL.
 CC CHAIN 1666 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
 CC CHAIN 1697 2470 NOTCH EXTRACELLULAR TRUNCATION.
 CC CHAIN 1697 2470 NOTCH INTRACELLULAR DOMAIN.
 CC DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 1678 1698 POTENTIAL.
 CC DOMAIN 1699 2470 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 26 63 EGF-LIKE 1.
 CC DOMAIN 64 102 EGF-LIKE 2.
 CC DOMAIN 105 143 EGF-LIKE 3.
 CC DOMAIN 144 180 EGF-LIKE 4.
 CC DOMAIN 182 219 EGF-LIKE 5. CALCIUM-BINDING (POTENTIAL).
 CC DOMAIN 221 256 EGF-LIKE 6 (INCOMPLETE).
 CC DOMAIN 258 294 EGF-LIKE 7. CALCIUM-BINDING (POTENTIAL).
 CC DOMAIN 296 334 EGF-LIKE 8. CALCIUM-BINDING (POTENTIAL).
 CC DOMAIN 336 372 EGF-LIKE 9. CALCIUM-BINDING (POTENTIAL).
 CC DOMAIN 373 411 EGF-LIKE 10.
 CC DOMAIN 413 452 CALCIUM-BINDING (POTENTIAL).
 CC DOMAIN 454 490 CALCIUM-BINDING (POTENTIAL).
 CC DOMAIN 492 528 EGF-LIKE 13. CALCIUM-BINDING (POTENTIAL).

FT DOMAIN	530	566	EGF-LIKE 14,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	568	603	EGF-LIKE 15,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	605	641	EGF-LIKE 16,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	643	678	EGF-LIKE 17,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	680	716	EGF-LIKE 18,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	718	753	EGF-LIKE 19,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	755	791	EGF-LIKE 20,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	793	829	EGF-LIKE 21,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	831	869	EGF-LIKE 22,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	871	907	EGF-LIKE 23,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	909	945	EGF-LIKE 24,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	947	983	EGF-LIKE 25,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	985	1021	EGF-LIKE 26,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	1023	1059	EGF-LIKE 27,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	1061	1097	EGF-LIKE 28,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	1099	1145	EGF-LIKE 29,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	1147	1183	EGF-LIKE 30,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	1185	1221	EGF-LIKE 31,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	1223	1260	EGF-LIKE 32,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	1262	1300	EGF-LIKE 33,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	1302	1345	EGF-LIKE 34,	CALCIUM-BINDING	(POTENTIAL)
FT DOMAIN	1372	1410	EGF-LIKE 35,	CALCIUM-BINDING	(POTENTIAL)
FT REPEAT	1418	1454	LIN/NOTCH 1,		
FT REPEAT	1501	1533	LIN/NOTCH 2,		
FT REPEAT	1825	1869	ANK 1,		

Query Match 1.1%; Score 8; DB 1; Length 2470;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 426 PCOHGATC 433
 Db 1155 PCOHGATC 1162

Search completed: March 9, 2004, 17:28:34
 Job time: 19.0517 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:22:43 ; Search time 61.7389 Seconds
(without alignments)
3602.917 Million cell updates/sec

Title: US-10-006-011A-3
Perfect score: 705
Sequence: 1 EIKIFRPSADGMLLYNQ.....QLDLQHRQAQANTRPCPS 705

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SPREMBL_25.*

- 1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp Vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Table with columns: Result No., Score, Query Match, Length, DB ID, Description. Lists various species and their corresponding sequence identifiers.

Table with columns: ID, Score, Query Match, Length, DB ID, Description. Lists various species and their corresponding sequence identifiers.

ALIGNMENTS

RESULT 1
O08591 ID O08591 PRELIMINARY; PRT; 68 AA.
AC O08591;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Parlecan (Fragment)
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=Kidney;
RA Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
RT "Proteoglycan expression in the normal rat kidney.";
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U75305; AAB51124.1;
DR InterPro; IPR008985; ConA_like lec_gl.
DR Pfam; PF00054; laminin_G_1.
DR PROSITE; PSS0025; LAM1_G_DOMAIN; 1.
FT NON_TER 1
FT NON_TER 68
SQ SEQUENCE 68 AA; 7395 MW; 5868E45DBA7083E0 CRC64;

Query Match 1.8%; Score 13; DB 11; Length 68;
Best Local Similarity 100.0%; Pred. No. 4.5e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 533 FRSRLPEVPEETIE 545
Db 21 FRSRLPEVPEETIE 33

RESULT 2
Q9DFLO ID Q9DFLO PRELIMINARY; PRT; 53 AA.

AC Q9DFLO:
 DC 01-WAR-2001 (TREMBlrel. 16, Created)
 DT 01-WAR-2001 (TREMBlrel. 16, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE S-adenosylmethionine synthase-like protein (Fragment)
 OS Gilllichthys mirabilis (long-jawed mudsucker)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Gobioidei;
 OC Gobiidae; Gilllichthys.
 CN NCBI_TaxID=8222;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA MEDLINE=21117151; PubMed=11172064;
 RA Gracey A.Y., Troll J.V., Somero G.N.;
 RT "Hypoxia-induced gene expression profiling in the euryoxic fish
 Gilllichthys mirabilis."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:1993-1998(2001).
 DR EMBL; AF266225; AAG13345.1; -
 FT NON_TER 1
 FT SEQUENCE 53 AA; 5541 MW; A7D6F0FP411CF46E CRC64;

Query Match 1.4%; Score 10; DB 13; Length 53;
 Best Local Similarity 100.0%; Pred. No. 0.053;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 QGTRCLCLPG 483
 Db 35 QGTRCLCLPG 44

RESULT 3
 ID Q80YAS PRELIMINARY; PRT; 1208 AA.
 AC Q80YAS
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DE 593040A21 protein (fragment).
 GN 593040A21.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CN NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuller G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heib H.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley D.M., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Raey J., Helton E., Kettner M., Madan A.C., Rodriguez S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Strausberg R.;

RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; BC043114; AAH43114.1; -
 DR GO; GO:000509; F:calcium ion binding; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2
 DR InterPro; IPR001881; EGF_CA.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR PRINTS; PR00008; EGF; 13.
 DR PRINTS; PR00010; EGFLOOD.
 DR PRINTS; PR00011; EGFLAMININ.
 DR SMART; SM00181; EGF; 15.
 DR SMART; SM00179; EGF_CA; 13.
 DR SMART; SM00282; LamG; 3.
 DR PROSITE; PS00010; ASX_HYDROXYL; 8.
 DR PROSITE; PS00022; EGF_1; 14.
 DR PROSITE; PS01186; EGF_2; 9.
 DR PROSITE; PS01187; EGF_CA; 5.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 2.
 FT NON_TER 1
 FT SEQUENCE 1208 AA; 126978 MW; FDF2DF2F0B2F198 CRC64;

Query Match 1.3%; Score 9; DB 11; Length 1208;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 EFLHGTC 473
 Db 295 EFLHGTC 303

RESULT 4
 ID Q7TQ52 PRELIMINARY; PRT; 2516 AA.
 AC Q7TQ52
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 GN Transmembrane receptor Notch1 B.
 DE NOTCH1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CN NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Thymus;
 RA Tsuji H., Ishii-Obba H., Ukai H., Katsube T., Ogiu T.;
 RT "Radiation-induced deletions in the 5' end region of Notch1 lead to
 the formation of truncated proteins and are involved in the
 development of mouse thymic lymphomas."
 RL Carcinogenesis 24:1-12(2003).
 DR EMBL; AB100603; BAC77038.1; -
 KW Receptor; Transmembrane
 SQ SEQUENCE 2516 AA; 269177 MW; 17FD72740EBD6E35 CRC64;

Query Match 1.3%; Score 9; DB 11; Length 2516;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCLHGTCQ 474
 Db 1014 PCLHGTCQ 1022

RESULT 5
 ID Q7TQ51 PRELIMINARY; PRT; 2526 AA.
 AC Q7TQ51

DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Transmembrane receptor Notch1 D.
 GN NOTCH1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C.B-17; TISSUE=Thymus;
 RA Tsuji H., Ishii-Ohba H., Ukai H., Katsube T., Ogiu T.;
 RT "Radiation-induced deletions in the 5' end region of Notch1 lead to
 RT the formation of truncated proteins and are involved in the
 RT development of mouse thymic lymphomas."
 RL Carcinogenesis 24:1-12(2003).
 DR EMBL; AB100603; BAC77039.1; --
 KW Receptor; Transmembrane.
 SQ SEQUENCE 2526 AA; 270583 MW; 017563FCE9703264 CRC64;

 Query Match 1.3%; Score 9; DB 11; Length 2526;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 466 PCLHGTCQ 474
 Db 1024 PCLHGTCQ 1032

 RESULT 6
 ID Q8K428 PRELIMINARY; PRT; 2531 AA.
 AC Q8K428
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Notch 1 protein.
 GN NOTCH1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Thymus;
 RX MEDLINE=95044925; PubMed=7956822;
 RA Nye J.S., Kopan R., Axel R.;
 RT "An activated Notch suppresses neurogenesis and myogenesis but not
 RT gliogenesis in mammalian cells."
 RL Development 120:2421-2430(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Thymus;
 RX MEDLINE=22119593; PubMed=12123574;
 RA Foltz D.R., Santiago M.C., Berechid B.E., Nye J.S.;
 RT "Glycogen Synthase Kinase-3beta Modulates Notch Signaling and
 RT Stability."
 RL Curr. Biol. 12:1006-1011(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Thymus;
 RA Schellin K.A., Paulley A.M., Nye J.S.;
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF508809; AAM28905.1; --
 DR MGD; MGI_97363; Notch1
 DR GO; GO:0005515; P:protein binding; IPI.
 DR GO; GO:0030194; P:cell differentiation; IMP.
 DR GO; GO:0007386; P:compartment specification; IMP.
 DR GO; GO:0045944; P:positive regulation of transcription from P. . .; IDA.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.

DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 36.
 DR Pfam; PF00066; notch; 3.
 DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR00011; EGFFLAMIN.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SMC0248; ANK; 6.
 DR SMART; SMC0181; EGF; 37.
 DR SMART; SMC0179; EGF_CA; 35.
 DR SMART; SMC0004; NL; 3.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 35.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS01187; EGF_CA; 21.
 DR PIRSF; PIRSF002279; Notch; 1.
 KW ANK repeat; EGF-like domain; Repeat.
 SQ SEQUENCE 2531 AA; 270819 MW; 7DB7E0DEF799D999 CRC64;

 Query Match 1.3%; Score 9; DB 11; Length 2531;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 466 PCLHGTCQ 474
 Db 1029 PCLHGTCQ 1037

 RESULT 7
 ID Q7TQ50 PRELIMINARY; PRT; 2531 AA.
 AC Q7TQ50
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Transmembrane receptor Notch1.
 GN NOTCH1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C.B-17; TISSUE=Thymus;
 RA Tsuji H., Ishii-Ohba H., Ukai H., Katsube T., Ogiu T.;
 RT "Radiation-induced deletions in the 5' end region of Notch1 lead to
 RT the formation of truncated proteins and are involved in the
 RT development of mouse thymic lymphomas."
 RL Carcinogenesis 24:1-12(2003).
 DR EMBL; AB100603; BAC77040.1; --
 KW Receptor; Transmembrane.
 SQ SEQUENCE 2531 AA; 270832 MW; 97C91F69BABF02BF CRC64;

 Query Match 1.3%; Score 9; DB 11; Length 2531;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 466 PCLHGTCQ 474
 Db 1029 PCLHGTCQ 1037

 RESULT 8
 ID Q9XZC9 PRELIMINARY; PRT; 3367 AA.

Q9XZC9; 01-NOV-1999 (T-EMBLrel. 12, Created)
 DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Laminin ALPHA1.2 (SYMBOL=WB)
 GN WB OR WING BLISTER OR CG15288.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=99403001; PubMed=10471707;
 RA Ashburner M., Misra S., Rote J., Lewis S.E., Blazej R., Davis T.,
 RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
 RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
 RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
 RA Celniker S., Rubin G.M.;
 RT "An exploration of the sequence of a 2.9-Mb region of the genome of
 RT Drosophila melanogaster: the Adh region.";
 RL Genes 153:179-219(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RA Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
 RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
 RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
 RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
 RA Lewis S., Li P., Lomontan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
 RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequelra A.,
 RA Sethi H., Snir E., Swirekas R.R., Wan K.H., Weinburg T., Zhang R.,
 RA Zieran L.L., Rubin G.M.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF135118; AAD31714.1;
 DR EMBL; AE003409; AAF44861.1;
 DR HSSP; P00740; 1EDM
 DR GO; GO:0007267; F:cell-cell signaling; NAS.
 DR InterPro; IPR001589; Actbind_actinin.
 DR InterPro; IPR008985; Cona_like_lec_gl.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR008211; LamNT.
 DR Pfam; PF00052; laminin_B; 2.
 DR Pfam; PF00053; laminin_EGF; 14.
 DR Pfam; PF00054; laminin_G; 4.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR SMART; SM00180; EGF_Lam; 16.
 DR SMART; SM00282; LamG; 5.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS00022; EGF_1; 14.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 16.
 DR PROSITE; PS50025; LAM_G_DOWNIN; 4.
 KW Laminin EGF-like domain.
 SQ SEQUENCE 3367 AA; 374097 MW; EB125654B1BC1511 CRC64;

Query Match 1.3%; Score 9; DB 5; Length 3367;

Best Local Similarity 100.0%; Pred. No. 26;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 603 NDGEWHRVT 611
 |||||
 Db 3081 NDGEWHRVT 3089
 RESULT 9
 Q8IP51 PRELIMINARY; PRT; 3375 AA.
 ID Q8IP51
 AC Q8IP51;
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DE 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE CG15288-PE.
 GN WB OR CG15288.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton R.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.G., Rogers J.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Chery J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Galbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskeir D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Follard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu S.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banjon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorset V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreriz S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA Mcintosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,
 RA Phouanavong S Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Swirskas K., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Bertram B., Carlson J.W., Ceiniker S.E.,
 RA Klomp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB003643; AAN10875.1; -;
 DR FlyBase; FBgn004002; wb.
 DR GO; GO:007267; P:cell-cell signaling; NAS.
 DR InterPro; IPR001589; Actbind actnin.
 DR InterPro; IPR009885; ConA_like_lec_gl.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; ISGF.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR008211; LamNT.
 DR Pfam; PF00052; laminin_B; 2.
 DR Pfam; PF00053; laminin_EGF; 14.
 DR Pfam; PF00054; laminin_G; 4.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_Laminin.
 DR SMART; SM00181; EGF; 9.
 DR SMART; SM00180; EGF_Lam; 17.
 DR SMART; SM00281; LamB; 2.
 DR SMART; SM00282; LamG; 5.
 DR SMART; SM00136; LamWT; 1.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS01186; EGF_1; 14.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 16.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 4.
 SQ SEQUENCE 3375 AA; 374742 MW; 38CB65C01BB6E416 CRC64;

Query Match 1.3%; Score 9; DB 5; Length 3375;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 603 NDGEWHRVT 611
 |||||
 Db 3089 NDGEWHRVT 3097
 RESULT 10
 Q81718 PRELIMINARY; PRT; 97 AA.
 AC Q81718
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE tRNA intron endonuclease, putative.
 GN Pf14_0514.
 OS Plasmodium falciparum (isolate 3D7).

OC Rukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=3D7;
 RX MEDLINE=22355705; PubMed=12368864;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
 RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
 RA McFadden G.I., Cummings L.W., Subramanian G.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrell B.;
 RT "Genome sequence of the human malaria parasite Plasmodium
 falciparum";
 RL Nature 419:498-511(2002).
 RL EMBL; AE014824; AAN37127.1; -;
 DR GO; GO:0000214; C:tRNA-intron endonuclease complex; IEA.
 DR GO; GO:0004519; F:endonuclease activity; IEA.
 DR GO; GO:0000213; F:tRNA-intron endonuclease activity; IEA.
 DR GO; GO:0006388; P:tRNA splicing; IEA.
 DR InterPro; IPR006677; tRNA_int_endo_C.
 DR Pfam; PF01974; tRNA_int_endo; 1.
 KW Endonuclease.
 SQ SEQUENCE 97 AA; 11640 MW; D27B07F8A1360441 CRC64;
 Query Match 1.1%; Score 8; DB 5; Length 97;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 143 EIVFHDLN 150
 |||||
 Db 9 EIVFHDLN 16
 RESULT 11
 Q8BM15 PRELIMINARY; PRT; 219 AA.
 AC Q8BM15
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23; Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25; Last annotation update)
 DE weakly similar to neurogenic locus notch 3 protein (Fragment).
 OS Mus musculus (Mouse).
 OC Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=C57BL/6J; TISSUE=Forelimb;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 DR EMBL; AK031068; BAC27237.1; -;
 DR FIR; FT0633; FT0633.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR InterPro; IPR000152; Aex hydroxyl_S.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR006210; IEGF.
 DR Pfam; PF00008; EGF; 2.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00179; EGF_CA; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; 4.
 DR PROSITE; PS01185; EGF_2; 2.
 DR NON_TER

SQ SEQUENCE 219 AA; 23098 MW; 9F2C4BE237F9B522 CRC64;

Query Match 1.1%; Score 8; DB 11; Length 219; Best Local Similarity 100.0%; Pred. No. 25; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 482 PGFSGPRC 489

DB 56 PGFSGPRC 63

RESULT 12

O9P6X9 ID Q9P6X9 PRELIMINARY; PRT; 244 AA. AC Q9P6X9; DT 01-OCT-2000 (TREMBlrel. 15, Created) DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update) DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update) DE Related to putative lipid binding protein TFS1. GN I3E11.370. OS Neurospora crassa. OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora. OX NCBI_TaxID=5141; RN [1] SEQUENCE FROM N.A. RP SEQUENCE FROM N.A. RA Schulte U., Aign V., Hobeisel J., Brandt P., Fartmann B., Holland R., RA Nyakatura G., Mewes H.W., Mannhaupt G.; RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases. RN [2] SEQUENCE FROM N.A. RP German Neurospora genome project; RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases. DR EMBL; AL353820; CAB98605.2; -. DR HSP; F30086; IBD9. DR InterPro; IPR008914; PEPF. DR Pfam; PF01161; PBP; 1. SQ SEQUENCE 244 AA; 26745 MW; EEB0DC9E6245012A CRC64;

Query Match 1.1%; Score 8; DB 3; Length 244; Best Local Similarity 100.0%; Pred. No. 27; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 90 LAPVNGTS 97

DB 186 LAPVNGTS 193

RESULT 13

Q82W41 ID Q82W41 PRELIMINARY; PRT; 302 AA. AC Q82W41; DT 01-JUN-2003 (TREMBlrel. 24, Created) DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update) DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update) DE Phosphadenosine phosphosulfate reductase (EC 2.7.7.4). GN CYSD OR NE0856. OS Nitrosomonas europaea. OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales; OC Nitrosomonadaceae; Nitrosomonas. OX NCBI_TaxID=915; RN [1] SEQUENCE FROM N.A. RP STRAIN=ATCC 19718 / IFO 14298; RX MEDLINE=22586410; PubMed=12700255; RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M., RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Savayendra-Soto L.A., RA Arclero D.M., Hommes N.G., Whittaker M.M., Atp D.J.; RT "Complete genome sequence of the ammonia-oxidizing bacterium and RT obligate chemolithoautotroph Nitrosomonas europaea."; RL J. Bacteriol. 185:2759-2773(2003). DR EMBL; BX321859; CAD84767.1; -. DR GO; GO:0016779; F:nucleotidyltransferase activity; IEA.

DR GO; GO:0004781; F:sulfate adenylyltransferase (ATP) activity; IEA. DR GO; GO:0016740; F:transferase activity; IEA. DR InterPro; IPR002500; PAPS_reduct. DR Pfam; PF01507; PAPS_reduct; 1. KW Nucleotidyltransferase; Transferase; Complete proteome. SQ SEQUENCE 302 AA; 34973 MW; 5499EAC71EF9F7DD CRC64;

Query Match 1.1%; Score 8; DB 16; Length 302; Best Local Similarity 100.0%; Pred. No. 33; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 279 VLLFSGGK 286

DB 32 VLLFSGGK 39

RESULT 14

Q842M1 ID Q842M1 PRELIMINARY; PRT; 308 AA. AC Q842M1; DT 01-JUN-2003 (TREMBlrel. 24, Created) DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update) DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update) DE CYSD (EC 2.7.7.4). GN CYSD. OS Rhodococcus sp. DS7. OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; OC Corynebacterineae; Nocardiaceae; Rhodococcus. OX NCBI_TaxID=227347; RN [1] SEQUENCE FROM N.A. RP STRAIN=DS7; RC Vitale E., Brunelli E., de Ferra F.; RT "Effects of cybNC mutations on the desulphurization of organosulphur compounds in Rhodococcus sp. strain DS7."; RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases. DR EMBL; AY244763; AAO89190.1; DR GO; GO:0016779; F:nucleotidyltransferase activity; IEA. DR GO; GO:0004781; F:sulfate adenylyltransferase (ATP) activity; IEA. DR GO; GO:0016740; F:transferase activity; IEA. DR InterPro; IPR002500; PAPS_reduct. DR Pfam; PF01507; PAPS_reduct; 1. KW Nucleotidyltransferase; Transferase. SQ SEQUENCE 308 AA; 35136 MW; AE64C09A9D453F7C CRC64;

Query Match 1.1%; Score 8; DB 2; Length 308; Best Local Similarity 100.0%; Pred. No. 34; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 279 VLLFSGGK 286

DB 37 VLLFSGGK 44

RESULT 15

Q8G4F5 ID Q8G4F5 PRELIMINARY; PRT; 320 AA. AC Q8G4F5; DT 01-MAR-2003 (TREMBlrel. 23, Created) DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update) DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update) DE Probable ribokinase. GN BL1431. OS Bifidobacterium longum. OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales; OC Bifidobacteriaceae; Bifidobacterium. OX NCBI_TaxID=216816; RN [1] SEQUENCE FROM N.A. RP STRAIN=NCC 2705; RX MEDLINE=22294977; PubMed=12381787;

RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
 RA Pessi G., Zwalhen M.-C., Desiere F., Bork P., Delley M.,
 RA Prigmore R.D., Arigoni F.,
 RT "The genome sequence of *Bifidobacterium longum* reflects its adaptation
 RT to the human gastrointestinal tract,"
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
 DR EMBL; AF014773; AAN25228.1; .
 DR GO; GO:0004747; F:ribokinase activity; IEA.
 DR GO; GO:0006014; P:D-ribose metabolism; IEA.
 DR InterPro; IPR002173; PfKB.
 DR InterPro; IPR002139; Ribokinase.
 DR Pfam; PF00294; PfKB; 1
 DR PRINTS; PRO0950; RIBOKINASE.
 DR PROSITE; PS00583; PFKB_KINASES_1; 1.
 KW Kinase; Complete proteome.
 SQ SEQUENCE 320 AA; 33130 MW; 28F44619036EFC6B CRC64; .

Query Match 1.1%; Score 8; DB 16; Length 320;
 Best Local Similarity 100.0%; Pred.No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 GRWHVSA 332
 Db |||||
 249 GRWHVSA 256

Search completed: March 9, 2004, 17:30:33
 Job time : 64.739 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:15:33 ; Search time 24.1076 Seconds
(without alignments)
2813.016 Million cell updates/sec

Title: US-10-006-011A-3
Perfect score: 3825
Sequence: 1 EKITFRPDSADGNLLYNGQ.....QLDLQHQAGANTRCPS 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.78:*
1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3825	100.0	4391	A38096	perlecan precursor
2	3389.5	88.6	3707	S18252	heparan sulfate pr
3	859	22.5	1959	AGRT	agrin - rat
4	796.5	20.8	1955	AGCH	agrin precursor -
5	667	17.4	1328	T43050	hypothetical prote
6	511	13.4	3375	T19821	neurexin II-alpha
7	421	11.0	1715	C40228	slit-1 protein hom
8	419.5	11.0	1531	T42218	neurexin I-alpha
9	405.5	10.6	1530	T45944	neurexin I-alpha p
10	403.5	10.5	1507	A40228	secreted leucine-r
11	387.5	10.1	1025	T42626	MEGFS protein - ra
12	387	10.1	1523	T13953	cadherin-related t
13	383	10.0	5147	IJUF7M	neurexin III-alpha
14	382	10.0	1438	A48216	neurexin III-alpha
15	382	10.0	1471	F48218	neurexin III-alpha
16	382	10.0	1578	T48216	hypothetical prote
17	362.5	9.5	4307	T20721	laminin alpha-2 ch
18	357	9.3	1751	MMHDMH	laminin alpha-1 ch
19	353.5	9.2	3084	MMMSA	laminin alpha-1 ch
20	353	9.2	3075	S14458	DN-cadherin - fru1
21	349	9.1	3097	T00021	laminin alpha-2 ch
22	347	9.1	3106	S53868	hypothetical prote
23	344.5	9.0	601	T22025	protein F40E10.4 [
24	344.5	9.0	601	D89711	notch homolog - se
25	340	8.9	2531	T31070	slit protein 2 pre
26	338.5	8.8	1469	B36665	slit protein 1 pre
27	333	8.7	1480	A36665	hypothetical prote
28	326	8.5	3672	T23433	cell-fate determin
29	319	8.3	2471	A49128	

30	317	8.3	3704	2	T37316	probable laminin a
31	316	8.3	2321	2	S78549	notch3 protein - h
32	314.5	8.2	1203	2	A49175	Notch B protein -
33	311	8.1	2318	2	S45306	notch 3 protein -
34	311	8.1	3034	2	T14119	seven-pass transme
35	304	7.9	2809	2	T30213	G-cadherin - sea u
36	302.5	7.9	1220	2	A56136	jagged protein pre
37	302.5	7.9	2555	2	A40043	notch protein homo
38	299.5	7.8	2524	2	A35844	Xotch protein - Af
39	298	7.8	4351	2	T00252	MEGFI protein - ra
40	296	7.7	2531	2	S18188	notch protein homo
41	296	7.7	2531	2	A46019	notch-1 protein -
42	292	7.6	861	2	A48825	Notch homolog Motc
43	291	7.6	1964	2	T09059	notch4 - mouse
44	287.5	7.5	2352	2	T30201	Notch homolog prot
45	284.5	7.4	2703	1	A24420	notch protein - fr

ALIGNMENTS

RESULT 1

A38096
perlecan precursor - human
N;Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate prote
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text change 05-Nov-1999
C;Accession: A38096; S19256; S77946; A41059; A40306; B33625; A36225; A41736
R;Murdock, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J. Biol. Chem. 267, 8544-8557, 1992
A;Title: Primary structure of the human heparan sulfate proteoglycan from basement membr
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A;Reference number: A38096; MUID:92235084; PMID:15669102
A;Accession: A38096
A;Molecule type: mRNA
A;Residues: 1-4391 <MUR>
A;Cross-references: GB:M85289; NID:g184426; PIDN:AAA52700.1; PID:g184427
R;Kallunki, P.; Tryggvason, K.
J. Cell Biol. 116, 559-571, 1992
A;Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD pro
tein adhesion molecules, and epidermal growth factor.
A;Reference number: A41736; MUID:92112994; PMID:1730768
A;Accession: S19256
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'E',438-449,'Q',451-502,'A',503-792,'K',794-908,'R',
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-3.
A;Cross-references: EMBL:X62515
R;Tryggvason, K.
submitted to the EMBL Data Library, October 1991
A;Reference number: S77946
A;Accession: S77946
A;Molecule type: mRNA
A;Residues: 1-57,'D',59-434,'A',436,'E',438-449,'Q',451-502,'A',503-792,'K',794-908,'R',
71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-4.
A;Cross-references: EMBL:X62515; NID:G29469; PIDN:CAA44373.1; PID:G29470
R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
Genomics 11, 369-396, 1991
A;Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the g
A;Reference number: A41059; MUID:92120660; PMID:1685141
A;Accession: A41059
A;Molecule type: mRNA
A;Residues: 'RT',892-908,'R',910-1101,'L',1103-1132,'L',1134-1221,'L',1223-1397 <KA2>
A;Cross-references: NID:G243370; PIDN:AA821121.1; PID:G243371
R;Dodge, G.R.; Kovalszky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo, J
Genomics 10, 673-680, 1991
A;Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellula
A;Reference number: A40306; MUID:91365376; PMID:1679749
A;Accession: A40306
A;Molecule type: mRNA
A;Residues: 1018-1405,'G',1407-1409,'G',1411-1465 <DOD>
A;Cross-references: GB:M64283; NID:g184424; PIDN:AA53699.1; PID:g184425
R;Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den
J. Cell Biol. 109, 3199-3211, 1989

A;Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclonal antibodies.
 A;Accession: A33625, MUID:90078352; PMID:2687294
 A;Molecule type: protein
 A;Residues: 1379-1384, 'X', 1386-1388, 'X', 1390-1398 <HE2>
 A;Accession: A33625
 A;Molecule type: protein
 A;Residues: 2166-2171, 'X', 2173-2175, 'X', 2177-2185 <HE3>
 C;Genetic: A;Note: peptide potentially matches four different regions of sequence shown
 A;Gene: GDB:HSPG2
 A;Cross-references: GDB:126372; OMIM:142461
 A;Map position: ip36.1-1p36.1
 C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repeats
 C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembrane
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-4391/Product: perlecan #status predicted <MAT>
 F:22-193/Domain: I <DOM1>
 F:194-530/Domain: II <DOM2>
 F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F:531-1676/Domain: III <DOM3>
 F:1159-1206/Domain: laminin-type EGF-like homology <LEG>
 F:1563-1610/Domain: laminin-type EGF-like homology <LEG7>
 F:1613-1668/Domain: laminin-type EGF-like homology <LEG8>
 F:1677-3686/Domain: IV <DOM4>
 F:2007-2034/Domain: transmembrane #status predicted <TRM>
 F:3687-4391/Domain: V <DOM5>
 F:3845-3880/Domain: EGF homology <EGF1>
 F:3888-3921/Domain: EGF homology <EGF>
 F:3953-4106/Domain: laminin G repeat homology <LG2>
 F:4147-4175/Domain: EGF homology <EGF2>
 F:4149-4151/Region: motor neuron attachment (L-R-E) motif
 F:4299-4301/Region: motor neuron attachment (L-R-E) motif
 F:65,71,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
 F:89,554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
 Query Match 100.0%; Score 3825; DB 2; Length 4391;
 Best Local Similarity 100.0%; Pred. No. 5,4e-237;
 Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 421 PCRQPCQHGATCMTAPAGEYFQCLCRDGFKGLDCEHEENPCQLRREPCLHGTCGTRCLC 480
 Db 4107 PCRQPCQHGATCMTAPAGEYFQCLCRDGFKGLDCEHEENPCQLRREPCLHGTCGTRCLC 4166
 Qy 481 LPGFSGPRCCQSSGHGIAESDWHLESGGNDAPGQYGAFFHDDGFLAFPHGHVFSRSLPEV 540
 Db 4167 LPGFSGPRCCQSSGHGIAESDWHLESGGNDAPGQYGAFFHDDGFLAFPHGHVFSRSLPEV 4226
 Qy 541 PETIELEVRTSTASGLLLWQGVVEGAGQKDFISLGLQDGHILVFRYQLGSGEARLVSD 600
 Db 4227 PETIELEVRTSTASGLLLWQGVVEGAGQKDFISLGLQDGHILVFRYQLGSGEARLVSD 4286
 Qy 601 PINDGEWHRTALREGRRSIQVDGRELVSGRSPGNVAVNAKSGVYIGGAPDVAITLGG 660
 Db 4287 PINDGEWHRTALREGRRSIQVDGRELVSGRSPGNVAVNAKSGVYIGGAPDVAITLGG 4346
 Qy 661 RFSGGITGCVKLVLSARPGAPPQPLDLQHRQAQAGANTRPCPS 705
 Db 4347 RFSGGITGCVKLVLSARPGAPPQPLDLQHRQAQAGANTRPCPS 4391

RESULT 2
 S18252
 heparan sulfate proteoglycan - mouse
 N;Alternate names: perlecan
 C;Species: Mus musculus (house mouse)
 C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
 C;Accession: S18252; A31917; E31917; S66460
 R;Noonan, D.M.; Fuller, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaki, M.; Yamada, Y.; Ha
 J. Biol. Chem. 266, 22939-22947, 1991
 A;Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteogly
 adhesion molecule.
 A;Reference number: S18252; MUID:92078153; PMID:1744087
 A;Accession: S18252
 A;Molecule type: mRNA
 A;Residues: 1-3707 <NOO>
 A;Cross-references: EMBL:M77174; NID:G200295; PIDN:AAA39911.1; PID:G200296
 R;Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogteli, G.; Sasaki, M.; Yamada, Y.; Hass
 J. Biol. Chem. 263, 16379-16387, 1988
 A;Title: Identification of cDNA clones encoding different domains of the basement membra
 A;Reference number: A92680; MUID:89034110; PMID:2972708
 A;Accession: A31917
 A;Molecule type: mRNA
 A;Residues: 940-1601 <NO2>
 A;Cross-references: GB:J04054; NID:G200252; PIDN:AAA39899.1; PID:G200253
 A;Accession: B31917
 A;Molecule type: mRNA
 A;Residues: 1870-2600 <NO3>
 A;Cross-references: GB:J04055; NID:G200300; PIDN:AAA39912.1; PID:G200301
 R;Schulze, B.; Mann, K.; Battistutta, R.; Wiedemann, H.; Timpl, R.
 Eur. J. Biochem. 231, 551-556, 1995
 A;Title: Structural properties of recombinant domain III-3 of perlecan containing a glob
 A;Reference number: S66460; MUID:95377282; PMID:7649154
 A;Accession: S66460
 A;Molecule type: protein
 A;Residues: 1272-1274, 'X', 1276, 'X', 1278-1279 <SCH>
 C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
 C;Keywords: glycoprotein
 F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F:764-811/Domain: laminin-type EGF-like homology <LEG>
 F:1159-1206/Domain: laminin-type EGF-like homology <LEG7>
 F:1563-1610/Domain: laminin-type EGF-like homology <LEG8>
 F:1613-1668/Domain: laminin-type EGF-like homology <EG7>
 F:1677-3686/Domain: laminin-type EGF-like homology <EG8>
 F:2270-3423/Domain: EGF homology <EGF>
 F:3464-3492/Domain: EGF homology <EGF7>
 F:1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predicted
 Query Match 88.6%; Score 3389.5; DB 2; Length 3707;
 Best Local Similarity 88.9%; Pred. No. 3.9e-209;

Matches 627; Conservative 33; Mismatches 40; Indels 5; Gaps 3;

QY 1 EIKITFRSDSADGMLLYNGQKRVGSPFNLANRQDDFISFLVCGRPEFRFDAGSGMATI 60
 Db 3008 EIKITFRSDSADGMLLYNGQKRVGSPFNLANRQDDFISFLVCGRPEFRFDAGSGMATI 3064

QY 61 RHPTPLALGHFHTVTLRSLTQSSLIQSLVGLDAPVAVNGTSGKFGQGLDNEELYLGGYDYG 120
 Db 3065 RHPTPLALGHFHTVTLRSLTQSSLIQSLVGLDAPVAVNGTSGKFGQGLDNEELYLGGYDYG 3124

QY 121 IPKAGLSGGFTGCVRRLRIQGEIVFHDNLNHTAHGISHCPTCRDRPCONGGQCHDSSESS 180
 Db 3125 IPKAGLSGGFTGCVRRLRIQGEIVFHDNLNHTAHGISHCPTCRDRPCONGGQCHDSSESS 3184

QY 181 YVVCVPCAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGTCRCHLGRSLRCBEGVT 240
 Db 3185 YVVCVPCAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGTCRCHLGRSLRCBEGVT 3243

QY 241 VTTSLGAGSYLALPALTNTHTHELRLDVEPKLAPDGVLLFSGGKGGPVEDFVSLAMVG 300
 Db 3244 VTTSLGAGSYLALPALTNTHTHELRLDVEPKLAPDGVLLFSGGKGGPVEDFVSLAMVG 3303

QY 301 GHLEFRYELGSLAVLSAEPLALGRHRSVABRLNDGSLRVNGRPLVLRSSPGKSOGL 360
 Db 3304 GHLEFRYELGSLAVLSAEPLALGRHRSVABRLNDGSLRVNGRPLVLRSSPGKSOGL 3363

QY 361 NMHTLLYGGVPEVPSVPAATMSAHRFCVGEVSVNGKRLDLYSLGSGGICQCYDSS 420
 Db 3364 NMHTLLYGGVPEVPSVPAATMSAHRFCVGEVSVNGKRLDLYSLGSGGICQCYDSS 3423

QY 421 PCERQPCOHGATCMPAGEYEFQCLCRGFGKDLCEHEENPCOLREBCLHGTCQTRCLC 480
 Db 3424 PCERQPCOHGATCMPAGEYEFQCLCRGFGKDLCEHEENPCOLREBCLHGTCQTRCLC 3483

QY 481 LPFGSPRCQGGHGAESDWHLEGGGNDAPQYQAYFHDGDLAFPGHVFSRSLPEV 540
 Db 3484 LPFGSPRCQGGHGAESDWHLEGGGNDAPQYQAYFHDGDLAFPGHVFSRSLPEV 3543

QY 541 PETIELVRSSTAGSILLWQVVEGAEQGDKDFISLGLQDHLVFRYOLGSGEARLVSE 600
 Db 3544 PETIELVRSSTAGSILLWQVVEGAEQGDKDFISLGLQDHLVFRYOLGSGEARLVSE 3602

QY 601 PINDGEWHRTALREGRRGSIQVNDGELVSGRPGPNVAVNAKGVSVVIGGAPDVALITGG 660
 Db 3603 PINDGEWHRTALREGRRGSIQVNDGELVSGRPGPNVAVNAKGVSVVIGGAPDVALITGG 3662

QY 661 RFSGGITGVNVLVHARGAPPDPLDLQHRQAQAGANTRPCPS 705
 Db 3663 RFSGGITGVNVLVHARGAPPDPLDLQHRQAQAGANTRPCPS 3707

RESULT 3

agrin - rat
 A:Species: Rattus norvegicus (Norway rat)
 C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
 C:Accession: JH0399; A38856
 R:Rupp, F.; Pavan, D.G.; Magill-Solc, C.; Cowan, D.M.; Scheller, R.H. Neuron 6, 811-823, 1991
 A:Title: Structure and expression of a rat agrin.
 A:Reference number: JH0399; MUID:91222570; PMID:1851019
 A:Accession: JH0399
 A:Molecule type: mRNA
 A:Residues: 1-1779;1799-1959 <RUP>
 A:Cross-references: GB:M64780; NID:G202798; PIDN:AAA40703.1; PID:G202800
 A:Experimental source: embryonic spinal cord
 A>Note: it is uncertain whether Met-1, Met-18, or Met-24 is the initiator.
 R:Rupp, F.; Oexcelik, T.; Linnai, M.; Peterson, K.; Francke, U.; Scheller, R. J. Neurosci. 12, 3535-3544, 1992
 A:Title: Structure and chromosomal localization of the mammalian agrin gene.
 A:Reference number: A38856; MUID:92407628; PMID:1326608
 A:Accession: A38856
 A:Molecule type: mRNA

A:Residues: 1780-1798 <RU2>
 A:Cross-references: GB:S4194
 C:Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine r ycholine receptor clustering activity.
 C:Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repea
 C:Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction

F:1-1959/Product: agrin, form 1 #status predicted <AG1>
 F:1-1797,1799-1959/Product: agrin, form 3 #status predicted <AG3>
 F:1-1779,1788-1959/Product: agrin, form 5 #status predicted <AG5>
 F:1-1143,1153-1959/Product: agrin, form 2 #status predicted <AG2>
 F:22-50/Region: hydrophobic
 F:88-137/Domain: Kazal proteinase inhibitor homology <KPI1>
 F:163-212/Domain: Kazal proteinase inhibitor homology <KPI2>
 F:236-284/Domain: Kazal proteinase inhibitor homology <KPI3>
 F:307-356/Domain: Kazal proteinase inhibitor homology <KPI4>
 F:381-429/Domain: Kazal proteinase inhibitor homology <KPI5>
 F:446-494/Domain: Kazal proteinase inhibitor homology <KPI6>
 F:511-559/Domain: Kazal proteinase inhibitor homology <KPI7>
 F:540-542/Region: motor neuron attachment (L-R-E) motif
 F:596-645/Domain: Kazal proteinase inhibitor homology <KPI8>
 F:688-739/Domain: laminin-type EGF-like homology <LE1>
 F:742-786/Domain: laminin-type EGF-like homology <LE2>
 F:814-864/Domain: Kazal proteinase inhibitor homology <KPI9>
 F:869-992/Region: serine/threonine-rich
 F:1084-1086/Region: motor neuron attachment (L-R-E) motif
 F:1147-1215/Region: serine/threonine-rich
 F:1224-1257/Domain: EGF homology <EG1>
 F:1287-1442/Domain: laminin G repeat homology <LG1>
 F:1444-1478/Domain: EGF homology <EG2>
 F:1483-1515/Domain: EGF homology <EG3>
 F:1555-1706/Domain: laminin G repeat homology <LG2>
 F:1713-1747/Domain: EGF homology <EG4>
 F:1807-1959/Domain: laminin G repeat homology <LG3>
 F:97-116,105-137,171-191,180-212,244-263,252,284,316-335,324-356,389-408,397-429,454-473
 -1476,1483-1494,1488-1504,1506-1515/Disulfide bonds: #scatuent predicted
 F:145,672,827,957/Binding site: carbohydrate (Asn) #status predicted

Query Match 22.5%; Score 859; DB 1; Length 1959;
 Best Local Similarity 31.5%; Pred. No. 4.6e-47;

Matches 241; Conservative 92; Mismatches 278; Indels 154; Gaps 23;

QY 2 IKITFRPSADGMLLYNGQKRVGSPFNLANRQDDFISFLVCGRPEFRFDAGSGMATI 61
 Db 1287 LALEFRALTEGLLYNGNA-----RGKDFLALALLDGRVQFRFDTSGSPAVIT 1335

QY 62 HPTPLALGHFHTVTLRSLTQSSLIQSLVGLDAPVAVNGTSGKFGQGLDNEELYLGGYDYG-- 119
 Db 1336 SLVPEVPCGRWHELSRHRWQGLSDVGETFVVGESPGTDLNLDNLYVGGIPEQVA 1395

QY 120 -AIPKAGLSGGFTGCVRRLRIQGEIVFHDNLNHTAHGISHC--FTCRDRPCONGGQC 173
 Db 1396 MYLDRTSVGLKGCIRMLDINNQQLESLDQRAAVQSSGVGEGCDHPCLPNPCHGGALC 1455

QY 174 HDSSESSVVCVPCAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGTCRCHLGRSGL 233
 Db 1456 QALEAGMFLCQCFRFRGPTCADEKS-PCQNPFCGHAAPCVLSSG-GAKCECPGHSST 1513

QY 234 RCEBEGVTTTPSLSGAG-----SYLALPALNTHTHEL-----RLDVEFKPLAPDG 278
 Db 1514 FCQ---TV----LETAGSRPEPLADFNGFSYLELKLHTFERDLGKMALEVMVFLARGPSG 1566

QY 279 VLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSLAVLSAEPLALGRHRSVABRLNDG 338
 Db 1567 LLLYNGQKDGKGFVSLALNHRHLEFCYDLGKGAAVLRSEKPEFALGTWVFLERNGRK 1626

QY 339 GSLRVNGRPLVLRSSPGKSOQ-----LNLHTLLYGGVPEVPSVPAATMSAHRFCVGE 393
 Db 1627 GALTQVGDGPRVLGESP-KSRKRVPTMLNKEPLYIGAPDPFSLARGAAVSSGSGVQL 1685

QY 394 VSVNGKRLDLYSLGSGGICQCYDSSFCERQPCOHGATCMPAGEYEFQCLCRGFGKDL 453
 Db 394 VSVNGKRLDLYSLGSGGICQCYDSSFCERQPCOHGATCMPAGEYEFQCLCRGFGKDL 453

Db 1686 VSLRQHQL-----LTOEHLVRAVDVSPFADHPCCTQA----- 1716
 QY 454 CEHEENPCOLRBPCLHGHTC---OQT-RCLCLPFGSPRCOQSGHGIAESDWHLEGGG 509
 Db 1717 -----LGNPCLNGSCVPREATYECLECGFGSLHCKEGL-----VERASVG 1757
 QY 510 NDAPGOYGAIFHDDGFLAPPGHVFRSILPEV-----PEPI----- 544
 Db 1758 -----DLETLAPDGRTYEIVLNAVIESELNTEIPAPETLDSRALFSEKALQSN 1805
 QY 545 --ELEVRTSTAGLLWQVVEAGAGKDFISLQGHVFRYOLGSGEARLVSEDPPI 602
 Db 1806 HPELSLRTEATQGLVW-----LGAERADYVWALAVDGHQLQSLVDLGSQPVLRSVTKV 1861
 QY 603 NDEGHRVTVLREGRRGSIQVDEBELVSGRSPGNVAVAKGVSIVIGAP--DVATLTGG 660
 Db 1862 NTRNRLIRARHREHREGSLQVNEAPVTGSSPLGATQDLDGALWGLQKLPVGOALPK 1921
 QY 661 RPSSTGTCVKNLVLHSRPGAPPPQPLDQHRAGANTPCPS 705
 Db 1922 AYGTGFGVCLRDVVVGH-----RQLHLLLEDAVTKPELRPCPT 1958

RESULT 4
AGCH

agrin precursor - chicken
 C,Species: Gallus gallus (chicken)
 C,Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 17-Nov-2000
 C,Accession: JH0591; A38857; B38857; I50692
 R,Tsism, K.W.K.; Ruegg, M.A.; Escher, G.; Kroeger, S.; McMahan, U.J.
 Neuron 8, 677-689, 1992
 A,Title: cDNA that encodes active agrin.
 A,Reference number: JH0591; UID:92232297; PMID:1314620
 A,Accession: JH0591
 A,Molecule type: mRNA
 A,Residues: 1-1955 <TSI>
 A,Cross-references: GB:M94271; NID:G211120; PID:AAA48585.1; PID:G211121
 R,Ruegg, M.A.; Tsism, K.W.K.; Horton, S.E.; Kroeger, S.; Escher, G.; McMahan, U.J.
 Neuron 8, 691-699, 1992
 A,Title: The agrin gene codes for a family of basal lamina proteins that differ in function
 A,Reference number: A38857; UID:92232298; PMID:1314621
 A,Contents: alternative splicing
 A,Accession: A38857
 A,Molecule type: mRNA
 A,Residues: 1132-1783; 1795-1955 <RU2>
 A,Cross-references: GB:M97371
 A,Accession: B38857
 A,Molecule type: mRNA
 A,Residues: 1221-1647; 1652-1783; 1794-1955 <RU3>
 A,Cross-references: GB:M97372
 A,Note: translation of the nucleotide sequence is not complete
 R,Thomas, W.S.; O'Dowd, D.K.; Smith, M.A.
 Dev. Biol. 158, 523-535, 1993
 A,Title: Developmental expression and alternative splicing of chick agrin RNA.
 A,Reference number: I50692; UID:93345745; PMID:18393816
 A,Accession: I50692
 A,Status: preliminary; translated from GB/EMBL/DBDJ
 A,Molecule type: mRNA
 A,Residues: 'SHLSNEIPA', 1784-1795 <THO>
 A,Cross-references: EMBL:U07271; NID:9459665; PIDN:AAA16788.1; PID:9459666
 C,Comment: This protein mediates the motor neuron-induced aggregation of acetylcholine receptor family; agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repeat
 C,Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repeat
 C,Keywords: alternative splicing; duplication; glycoprotein; neuromuscular junction
 F,1-38/Domain: signal sequence #status predicted <MAT>
 F,39-1955/Product: agrin #status predicted <AG1>
 F,39-1783,1795-1955/Product: agrin-related protein 1 #status predicted <AG2>
 F,39-1647,1652-1783,1794-1955/Product: agrin-related protein 2 #status predicted <AG3>
 F,77-126/Domain: Kazal proteinase inhibitor homology <KPI1>
 F,152-201/Domain: Kazal proteinase inhibitor homology <KPI2>
 F,225-273/Domain: Kazal proteinase inhibitor homology <KPI3>
 F,295-344/Domain: Kazal proteinase inhibitor homology <KPI4>

F;370-418/Domain: Kazal proteinase inhibitor homology <KPI5>
 F;435-483/Domain: Kazal proteinase inhibitor homology <KPI6>
 F;500-548/Domain: Kazal proteinase inhibitor homology <KPI7>
 F;584-633/Domain: Kazal proteinase inhibitor homology <KPI8>
 F;675-726/Domain: laminin-type EGF-like homology <LEI1>
 F;729-773/Domain: laminin-type EGF-like homology <LEI2>
 F;801-851/Domain: Kazal proteinase inhibitor homology <KPI9>
 F;856-995/Region: serine/threonine-rich
 F;1150-1219/Region: serine/threonine-rich
 F;1233-1264/Domain: EGF homology <EG1>
 F;1294-1448/Domain: laminin G repeat homology <LGI1>
 F;1429-1431/Region: motor neuron attachment (L-R-E) motif
 F;1450-1482/Domain: EGF homology <EG2>
 F;1489-1521/Domain: EGF homology <EG3>
 F;1718-1751/Domain: laminin G repeat homology <LGI2>
 F;1803-1955/Domain: laminin G repeat homology <LGI3>
 F;185-105,94-126,160-180,189-201,233-252,241-273,304-323,312-344,378-397,386-418,443-462,
 1482,1489-1500,1494-1510,1512-1521/Disulfide bonds: #status predicted
 F;390,659,764,814/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 20.8%; Score 796.5; DB 1; Length 1955;
 Best Local Similarity 30.6%; Pred. No. 4,7e-43;
 Matches 221; Conservative 99; Mismatches 271; Indels 131; Gaps 25;

QY 2 IKITFRPDSADGMLLYNQKRVPGSPTNLANKQDFISFLVGGRRPFRFDAGSGMATIR 61
 Db 1294 IAMEFRATELSGLLYNQ-----NRKDFISLALVGGFVLFNFNFNGSGTGVIT 1342
 QY 62 HPTFLALGHFHVVTLRSLTQGLVLDLAPVNGYSQGFQGLDNEELYLGGYP-DYGA 120
 Db 1343 SKRVVPEGRKHQIVVNRRRSGLAV-DGEHVSQSGPTGDLNLDLDTLFGGAPEDQWA 1401
 QY 121 I--PKAGLSGGFIGVVR-----ELRIGQEEIVFHDLNLTARHGISHC--PTCRDRPC 167
 Db 1402 VVAERTAAATVGLKGSIRLLDVNNQVYDLREKGSVDLY-----GSGVGCNGDPCHPNPC 1455
 QY 168 QNGGQCHDESSVYCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGGYTCRH 227
 Db 1456 HFGASCHVKAEMFHCCELSHYTGPTCA-DERNPCDPTTCHISATCLVLIPEG-GAMCACP 1513
 QY 228 LGRSLRCEEGYVTVT-----TFLSAGSALPALPALTN---THHEIRLDVDFKPLAP 276
 Db 1514 MCRGEFCEP---VTEQDHTMPFLPEFNGF-SYLENLGLQTLFLTCRQMSMEVFLAKSF 1569
 QY 277 DGVLLPSSGKSPVEDFVSLAMVGGHLEFRYELGSLAVLSAEFLALGRWHRVSAERLN 336
 Db 1570 SGMIFYNGQKTDGKDFVSLALHDGYLEYRYDLGKGAVALRSKEFVPLNTWISVLLERSG 1629
 QY 337 KDGLRVNGGRVPLRASSPGKSGQ-----LNLHTLLYLGVEPSVPLSPATNNSAHRFGCV 391
 Db 1630 RKGVMRIINNGERVMSGESP-KSRKVFHAFNLKPEFFYVGGAPDFSKLARAALSTFSYGAV 1688
 QY 392 GEVSVNGKRLDLYTFLSGQGTGQCYSDSPCRQPCQHGATCMPAGEYEFQCLCRDQPKG 451
 Db 1689 QRISIKGVPL-----LKEQHRSVAEIST-----PRA 1715
 QY 452 DLCEHEENPCOLRBPCLHGHTC---OQTRCLCLPFGSPRCOQ-----GSGHGIA-- 498
 Db 1716 HFTOKPNEPCQ-----NGGTCSPLESEYECACRQFGSAGEKVIIEKAAGDAEIAFD 1769
 QY 499 ---ESDWHLEGGGNDAPCGYAYPHDDGFLAPGHVFRSRLPEVPEPIELEVRTSTASG 555
 Db 1770 GRTYMEYHNAVTKSPDA-----LDYPAAPSEKALQS--NHFELSITKEATQG 1814
 QY 556 LLLWQVVEGAGCGKDFISLQGHVFRYOLGSGEARLVSEDPINDGSHRVTALRE 615
 Db 1815 LILWS-----GKLESDVIALAVDGFQVMYDLSGKSPVLRSTVPIINTEHWHIKAYRV 1870
 QY 616 GRGSIQVDEBELVSGRSPGNVAVAKGVSIVIGAP--DVATLTGGFRSSGTCVKNL 673
 Db 1871 QREGSLQVNEAPITGSSPLGATQDLDGALWGLQKLPKAYS*GFIGCIRDV 1930

QY 674 VL 675
DB 1931 IV 1932

RESULT 5

T43060
agrin - electric ray (Discopyge ommata) (fragment)
C:Species: Discopyge ommata
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 17-Nov-2000
C/Accession: T43060
R:Smith, M.A.; Magill-Solc, C.; Rupp, F.; Yao, Y.M.M.; Schilling, J.W.; Snow, P.; McMahana
submitted to the EMBL Data Library, September 1992
A/Reference number: Z22308
A/Accession: T43060
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: mRNA
A/Residues: 1-1328 <EMBI>
A/Cross-references: EMBL:L01423; NID:G213102; PID:G213103; PIDN:AAA49224.1
C/Superfamily: agrin; EGF homology; Kazal proteinase inhibitor homology; laminin G repeat
C/Keywords: glycoprotein; neuromuscular junction

Query Match 17.4%; Score 667; DB 2; Length 1328;
Best Local Similarity 27.0%; Pred. No. 6.2e-35;
Matches 200; Conservative 112; Mismatches 304; Indels 126; Gaps 24;

- QY 2 IKITFRPDSADGMLLYNGQKRVGSPNTLANRQPEISFOLVGRPEFRFDAGSGMATIR 61
Db 673 ISMEFRASNLDELPLVQWTEKGGK---LHFYRPS-----EGYVELRPNHGVWDGVT 721
QY 62 HPTPLALGHHTVLLRSLTQGLIIVGDLAPVNGTSGKQKFGQLDNEEYLYGGYDYG-- 119
Db 722 SKTLIRKPNWHYVGNRRRSGMLSYDGEFHLIGESPCTDGLNLDLDFLGSTPEDEMT 781
QY 120 -AIPKAGLSGFTGCVRELRIOEETVFDLNL-----LTAHGISHCPT--CRDRPCQMG 170
Db 782 LVTERITATKLGQICRLDDVNN---LIYDLQERSNDVLYGSGVGCNPNPCSPAPKRR 838
QY 171 GQCHDESSYVCVCPAGFTGSRK--EHSQALHCHPEACGPDATCVNRPDGRGYTCRCHL 228
Db 839 GKCHKMKAEMFHCSEYGFSPCTCADKHP--CDNPNCHQGANCVLPEG--GSKCECPM 894
QY 229 GRSGLRCE-----SGVTVTPSLSGAGSYLALPAL----TNTWHELELDVYFPELAPD 277
Db 895 GREGELCERVSEAEQDQKAFPEFNGL--SYLEMNGIHTFVSDLLQKLSMEVIFLAKDPN 953
QY 278 GVLLFSGSGSPVEDFVSLAVMGHLEFRYELGSLAVLRSAPLALGRWHRVSAERLTK 337
Db 954 GMFYNGQKTDGRGDFVSLNLRDGYLEFYDLGKGAVALRSKAPIPLNVMVTVYRNGR 1013
QY 338 DGSRLVN-----GGRPVLRSPPKSGQLNHLTLVLYGVEPSPVLSPATNMSAHRFCV 391
Db 1014 KGLMINKDELVSAGESPKSRKAP--HTALNMLKEAFYVGGAPDFNKFARAAGIISGFTGAI 1071
QY 392 GYVNVGKRLDLYTSLGQIQCVSDSPCRPQCHGATCMPAGEVEYFOCLCRDGFKG 451
Db 1072 QKLSLK-----SIFLLKKNENRNAMEIS-----NFRW 1098
QY 452 DLCEHENPCQLREPLCHGGTC----QGTRCLCLPFGSPRCQOQSGHGEIABSDWHLRGS 507
Db 1099 HACTKTRNPCQ-----NGGVCSPRLREYDCMCQRFSGFQCEKA-----LB-- 1139
QY 508 GGNDAFQYCAVPHDDGELAFGHVFSRSLPVPETIELEVETSTASGLLMLQGVVESEA 567
Db 1140 -EKSASGSSEVAFNGRTFYEYHNVTYRSEKAVQVNFEMSIKTEATKGLILMS-----GKI 1194
QY 568 GQKDFPISLQDGHVLFYIQLGSEARLVSDFINDGEWHRVTALEGRGRGSIQVQGE 627
Db 1195 AEKSDYIALAVDGFQVMTYDLGSKPVTLRSTIPVNTQWVAIKANRHHGYGLQVGN 1254
QY 628 LYSGRSPGNVAVNAGSVYIGAPPVATLTGR-----FSSGITCVKMLVLSRPPAP 683
Db 1255 PVTGSSFFAATQDLDGALMLGIEKLA--PGNRLPFAKYSTGTFIGCIKDVVIDR----- 1306

RESULT 6

T19821
hypothetical protein ZC101.2e - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 23-Sep-2002
C/Accession: T19821; T19819; T19820; T27488; T27489; T27487; A47648; B47648; C47;
R:Baynes, C.
submitted to the EMBL Data Library, March 1997
A/Reference number: Z19182
A/Accession: T19821
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-3375 <WIL>
A/Cross-references: EMBL:Z93375; PIDN:CA807569.1; GSPDB:GN00020; CESP:ZC101.2e
A/Experimental source: clone C38C6
A/Accession: T19819
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-2441,'R','KRKH','3369','GN','3372-3373','G','3375','RLRHRRNAQNGPLSRKTRTTTKLFGSW
A/Cross-references: EMBL:Z93375; PIDN:CA807567.1; GSPDB:GN00020; CESP:ZC101.2a
A/Experimental source: clone C38C6
A/Accession: T19820
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-1694,'H','1883-2441','R','KRKH','3369','GN','3372-3373','G','3375','RLRHRRNAQNGPL
A/Cross-references: EMBL:Z93375; PIDN:CA807568.1; GSPDB:GN00020; CESP:ZC101.2c
A/Experimental source: clone C38C6
R:Percy, C.
submitted to the EMBL Data Library, March 1997
A/Reference number: Z20375
A/Accession: T27490
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-3375 <WI2>
A/Cross-references: EMBL:Z93395; PIDN:CA807708.1; GSPDB:GN00020; CESP:ZC101.2e
A/Experimental source: clone ZC101
A/Accession: T27488
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-2441,'R','KRKH','3369','GN','3372-3373','G','3375','RLRHRRNAQNGPLSRKTRTTTKLFGSW
A/Cross-references: EMBL:Z93395; PIDN:CA807706.1; GSPDB:GN00020; CESP:ZC101.2a
A/Experimental source: clone ZC101
A/Accession: T27489
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-1128,'N','1299','DFARNSPS','1299','NNS','1303-1304','R','RHR','1544-1545','RIRVRS',155:
A/Cross-references: EMBL:Z93395; PIDN:CA807704.1; GSPDB:GN00020; CESP:ZC101.2b
A/Experimental source: clone ZC101
R:Rogalaki, T.M.; Williams, B.D.; Mullen, G.P.; Moerman, D.G.
Genes Dev 7, 1471-1484, 1993
A/Title: Products of the unc-52 gene in Caenorhabditis elegans are homologous to the cori
A/Reference number: A47648; MUID:93339574; PMID:8393416
A/Accession: A47648
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-546,'P','548-2441','R','KRKH','3369','GN','3372-3373','G','3375','RLRHRRNAQNGPLSR
A/Accession: B47648
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-546,'P','548-2198','D','2290','NAR','2294','L','2296','WHATE',2302-2303,'V',2305,'I

QY 684 PFQPLDLQHRACAGANTRPCFS 705
DB 1307 --QELQVEDALNNPTILRCEA 1326

218 -----DGRGYTRCHLG-----RSLRCERGVVTTTPSLGAGS 251
 874 ISVVPNFISLHSLVNGPQYDQCKDGIITTCYELNARFGLBAIVADVPYFKSRS---S 930
 252 YLALPALTNTHEHLRLDVEKPLAPDGLVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGS 311
 931 YLALATL-QAYASMHLPFPKTTAPDGLLNFNSGNG---NDFIVIELVKGVIHYVFDLGN 986
 312 GLAVLR--SAEPLALGRWHRSABR-LNKDGSRLVNGRPRVLRSSPKSGOGLNHLTLYL 368
 987 SPSLMKNSDKPYNDQNHVNVSRDPPGNVHTLKID-SRTVTHQSNQ-ARNLDLKGELYI 1044
 369 GVEPSPVLSPATNMSAH---FRGCGEVSNGKRLDLYSFLGSGIGCCY-----DSS 420
 1045 GGLSKNM-FSNPKLVASRDGFOGLASVDLNGRLPDLADAL--HRIGQVERGCDGPGST 1101
 421 PCRQPCQHGATCWPAGEYEFQCLCRDGFKGDLCHEENPCQLRPFCLHGGCCQTRCLC 480
 1102 TCTEESCANQVCLQQW-----DGF-----TDCDF----- 1126
 481 LPGFSGPRCOQ-----GSGHGIAESDWHLEGGNDAPGQYGFHDDGFLAFPGHYF 533
 1127 MTSYGGPVCNDPGTYTFYFGGALITVTM---PNDRPPSTR---MDRLAVGFSTHOR 1177
 534 SRSLPVEPIEIEVTRSTASGILLWQGVGEAGQGGKDFISLGLQDGHVLFYRQLGSGE 593
 1178 SAVL-----VRVDSASGL-----GDYLOLHIDQGVTVGVIFNVGTD 1213
 594 ARLVSEDP---INDGEWHRTALREGRRSIQVD---GRELVSGRSPGNVAV----- 640
 1214 --ITIDEPNAIVSDGKHYVFRFRSGNATLQVDSWPNERYPAGNFNRLAIAQRIP 1271
 641 -----NAKGSVYIGGADPVAITLGGFRSSGITCCVKN 672
 1272 YRLGRVVDEWLLDKGRQLTIFNSQAAIKIGGRDQ-----GRPQGVSGLYYN 1319

RESULT 8
 T42218
 slit-1 protein homolog - rat
 N;Alternate names: MEGF4 protein
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Aug-2002
 C;Accession: T42218
 R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
 Genomics 51, 27-34, 1998
 A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
 A;Reference number: Z14126; MUID:98360089; PMID:9693030
 A;Accession: T42218
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: mRNA
 A;Residues: 1-1531 <NAK>
 A;Cross-references: EMBL:AB011530; NID:G3449289; PIDN:BA032460.1; PID:G3449290
 A;Experimental source: strain Sprague-Dawley; Drain
 C;Genetics:
 C;Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein 1

Query Match 11.0%; Score 419.5; DB 2; Length 1531;
 Best Local Similarity 23.3%; Pred. No. 5.5e-19;
 Matches 173; Conservative 78; Mismatches 245; Indels 245; Gaps 32;

25 GSPNTLANRQDFISFLVGGP-PEFRF-----DAGSGMATHRHTPTLALGH 70
 758 GIPKAVTELYLDGNOFTLVPGQLSTFKYLQVLDLSNNKISSLSNFTNMSQLTILSY 817
 71 FHTVTL-----LRSTQGLIVGLDLPVNGTISQKFOGLDNEELYLGGYDVAIFK 123
 818 NALQCIPLPAFQGLRSURLSLHGND---VSTIQEIGFADVTSLSHLAIGANPLYCDCHL 874
 124 AGLSS-----GFTICVRELIQGEIIVFDLMLTAHG-----ISHCPT 161
 875 RWLSSWVKTYGEGIARCAGPPEMEGK-----LLLTTPAKKFECCGPPSLAVQAQKDP 928

162 CRDRPCQNGOCHDSESSYVVCVPCAGFTGSRCEHSQALHCHPEACPDATC-VNRPDGR 220
 929 CLSSPCQNGGTRCHNDPLVTRCTCPSGFKGRNCEVSLD-SCSSNPCCNGGTCHAQEGEDA 987
 221 GYTCRCHLGRSLRCERGVVTTTPSLGAGSYLALPALNTNHEHLRLDVEKPLAPDGLV 280
 988 GFTCSFSGFEGLTG----- 1002
 281 LPSGGKSGPVEDFVSLAMVGGHLEFRYELGSLAVLRSABFLALGRWHRSABRLANKDGS 340
 1003 -----GNWTDCC 1009
 341 LR---VNGREVLRSRSPKSOGLNLHT---LLYLIG-GVEPSPVLS-PATNMSAHFRGCY 391
 1010 VKHDCVANGVCV-----DGIQNYTCQCPLOYTRACEQLVDFCSFDLNPCCHEAQCV 1061
 392 GEVSVNGKXLDLTYSLGSGIGOC-YDSSPCERQPCOHGATCWPAGEYEFQCLCRDGFK 450
 1062 G--TPGPRCEVCPGYTGD-----NCSKNQDDCKDHQCCONGAACVDEIN-STACLCAGEYS 1114
 451 GDLCCE--HEENPCQLRPEPCLHGQTC--QGTR--CLCLPGFSGPRCOQSGHGAESDWH 503
 1115 GQLCEIIPAPRNSCEGTE-CQGANCAVDQSSRPVCQCLPGFSGPECEKLLLSVNFVDRD 1173
 504 LEGSGNDAPCQYGAIFHDDGFLAFPGHYVRSRSLPEVPE-TIELEVTRSTASGILLWQGV 562
 1174 LQ-----FTD-----LQNWPRANITLQVSTAEQDNGILLYNG- 1204
 563 EVGAGQKDFISLGLQDGHVLFYRQLGS-GEARLVSEDPINDGEWHRTALREGRRSI 621
 1205 -----DNDHIAVELYQGHVRSYDPSYPSAIYSAEITINDQGFHTVELVTFDQWVNL 1257
 622 QVDGEELVSGRSPGNVAVNAKGSVYIGGAP-DVAT-----LTGGRFSSGITCCVKN 672
 1258 SIDGSPMTMDNFQKHYTLNSEAPLYVGGMPVDVNSAAFRMLQWILNGTSPH----GCIRN 1313

673 LVLHES-----ARPGAPP 684
 1314 LYINNELODFTKOMKFGVVP 1334

RESULT 9
 I45944
 neurexin I-alpha - bovine
 C;Species: Bos primigenius taurus (cattle)
 C;Date: 19-Dec-1997 #sequence_revision 19-Dec-1997 #text_change 09-Dec-2002
 R;Ullrich, B.; Ushkaryov, Y.A.; Sudhof, T.C.
 Neuron 14, 497-507, 1995
 A;Title: Cartography of neurexins: more than 1000 isoforms generated by alternative spli.
 A;Reference number: I45944; MUID:95209856; PMID:7695896
 A;Accession: I45944
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: mRNA
 A;Residues: 1-1530 <ULL>
 A;Cross-references: GB:L14855; NID:G388560; PIDN:AAA74123.1; PID:G388561
 C;Superfamily: neurexin; EGF homology
 F;703-735/Domain: EGF homology <EGF>

Query Match 10.6%; Score 405.5; DB 2; Length 1530;
 Best Local Similarity 20.5%; Pred. No. 4.4e-18;
 Matches 203; Conservative 112; Mismatches 285; Indels 391; Gaps 42;

2 IKITRPPDSADGMLLY-NGQRVFGSPNTLANRQDFISFLVGGPPEFRFDAGSGMATI 60
 527 ISFDPRITTEPNGLILFSGKPRHQKAKHPQMKVDFFAIEMLDGHLVLLLDMSGGTIKI 586
 61 RHPTPLAL-----GHFHTVLLRSLTQGLIVGLDLPVNGTISQKFOGLDNEELYL 113
 587 K-----ALQKVNDEGWYHVDQRDGRSGTISVNTLRTPY--TAPGSEQLDDDELYLG 639
 114 GVPDYGAIPKAG-----LSSGFICVRELIQGEIIVFDLMLTAHGSHC 159

Db 640 GLPE---NKAGLVPTTEVWTTALNNGYVGCIRDLFDIGQSKDIRQMAEVQSTAGVKPSC 695
 QY 160 ---PTCRDRPCONGGCHDSESSYVCP-AGFTGSRCE----- 195
 Db 696 SREYAKPCLSNFKKNGMCRDG-WNRVYVDCSCTGYLGRSCEAREATVLSYDGSFMFKIQL 754
 QY 196 ---HSQA-----LHCHPEAC-- 207
 Db 755 FVVMHTAEADVLSRFRSQRAYGILMATTSRDSADTLRLLELDAGRVKLTVNLDCIRINCNS 814
 QY 208 --GPDATCV--NRPDGRGYTCR----- 225
 Db 815 SKGPETLFCAGYNLNDNWHVTRVVRGSKSLKLVDDQQAMTGMAGDHTRLEFHNIEGTI 874
 QY 226 -----CHL-GRSGLR--CEBGYVTVT 243
 Db 875 ITERRYLSSVPSNFIHQLSLTFNGMAYIDLCKNGIDIDYCELNARFGRFNIIADPVTFKT 934
 QY 244 PLSGAGSYLALPALTNTTHELRDLVEFKPLAPDQVLLFSGKSPVEDFVSLAMVGGHL 303
 Db 935 KS-----SYVALATL-QAYTSMHLFFQFKTSLDGLILYN--SGDGNDFIVVELVKGYL 985
 QY 304 EFRYELGSLAVLR--SAPLALGRWHRVSAER-----LNKDGSLRVNGRPRVLS 352
 Db 986 HVYFDLGNANLIKGSNKPLNDNQHNVMISRDTSNHLTKVTKITQITAG----- 1039
 QY 353 SPKSGOGLNHLTLVLLGGV--EPSVPLSPATNMSAHRFGCVGEVYVNGKRLDLYSFLGS 410
 Db 1040 ---ARNLDLKSPLDLYIGVAKETYSKPLVHAKEGFGCLASVDLNGRLPDL----- 1088
 QY 411 QGIGQCYDSSPCERQPCQHGATCMPAGEYFQCLCRDGFKGLDCEHENPCQLREPCPLHG 470
 Db 1089 ---ISDALFCN-----CGQIE-----RG--CEGPTTTCQ--EDSCSNQ 1118
 QY 471 GTC----QGTRCLC-LPGFSGPRCQO-----GSGHGIAESDWHLEGGNDAPQGYGA 518
 Db 1119 GVCLOQWDSICDCSMTSFGPLCNDPDTGTYTFSKGGQITVKK-----PNDRPPSTRA- 1172
 QY 519 YFHDGFLAFPGHVRFSRSLPEVETIELE---VRTSTASGLLLMQGVEVGBAGQKDFIS 575
 Db 1173 ---DRLAIGF-----STVQKEAVLVRVDSSSGL-----GDYLE 1202
 QY 576 IGLQDGHVFRYQLGSGEARLVSEDP-INDGEWHRVYVTRALREGRGSIQVDG---BELVS 630
 Db 1203 LHHQKIGKVFNGVTDDIAIESNAIINDGKYHVVRFRSGGNATLQVDSWPVIERYP 1262
 QY 631 GRSPGNVAV-----NAKGSVYIGG----- 650
 Db 1263 GNDNERLAIARQIPYRLGRVVDWMLLDKGRQLTFNSQATIIIGKKEQGFQGLSG 1322
 QY 651 -----APDVALTGRFSSGITGVKMLV 674
 Db 1323 LYVNLGLVNLMAENDANITAVGNVRLVGEVPSMTESTATAMQSEMSTIMETTTLA 1382
 QY 675 LHSARPGAPPPOPLDQHQRAQAGANTRPCPS 705
 Db 1383 TSTARREKPTKPEVSVQTTDDILVSAECP 1413

A;Residues: 1-1507 <USH>
 A;Cross-references: GB:M96374; NID:G205710; PIDN:AAA41704.1; PID:G205711
 C;Superfamily: neurexin; EGF homology
 C;Keywords: alternative splicing; transmembrane protein
 F;1-30/Domain: signal sequence #status predicted <SIG>
 F;31-1507/Domain: neurexin I-alpha #status predicted <MAT>
 F;680-712/Domain: EGF homology <EGF>
 F;1087-1119/Domain: EGF homology <EGF1>

Query Match 10.5%; Score 403.5; DB 2; Length 1507;
 Best Local Similarity 20.4%; Pred. No. 5.8e-18;
 Matches 202; Conservative 113; Mismatches 285; Indels 391; Gaps 42;

QY 2 IKITRPPDSADQMLLY-NGQKRVPGSPNTLANRQDFISFGLVGRGPRFRFDAGSGMATI 60
 Db 504 ISFDFTTPEPGLLILFESHGKPRHOKDAKHPQMKVYFFFAIEMLDGHLVLLDMSGSIKI 563
 QY 61 RHPTPAL-----GHFHTVLLRSLTQSGLLVGLD-APVNGTSGOKFCGLDLNLELYIG 113
 Db 564 K-----ALQKYNDEGWVHVDQDRGSRGTISVNTLRTFY--TAFGESEILDLDDELYIG 616
 QY 114 GVPDYCAIPKAG-----LSSGFICVRELRIOGEEIVFH---DLNLTAGHISHC 159
 Db 617 GLPE---NKAGLVPTTEVWTTALNNGYVGCIRDLFDIGQSKDIRQMAEIQSTAGVKPSC 672
 QY 160 ---PTCRDRPCONGGCHDSESSYVCP-AGFTGSRCE----- 195
 Db 673 SREYAKPCLSNFKKNGMCRDG-WNRVYVDCSCTGYLGRSCEAREATVLSYDGSFMFKIQL 731
 QY 196 ---HSQA-----LHCHPEAC-- 207
 Db 732 FVVMHTAEADVLSRFRSQRAYGILMATTSRDSADTLRLLELDAGRVKLTVNLDCIRINCNS 791
 QY 208 --GPDATCV--NRPDGRGYTCR----- 225
 Db 792 SKGPETLFCAGYNLNDNWHVTRVVRGSKSLKLVDDQQAMTGMAGDHTRLEFHNIEGTI 851
 QY 226 -----CHL-GRSGLR--CEBGYVTVT 243
 Db 852 ITERRYLSSVPSNFIHQLSLTFNGMAYIDLCKNGIDIDYCELNARFGRFNIIADPVTFKT 911
 QY 244 PLSGAGSYLALPALTNTTHELRDLVEFKPLAPDQVLLFSGKSPVEDFVSLAMVGGHL 303
 Db 912 KS-----SYVALATL-QAYTSMHLFFQFKTSLDGLILYN--SGDGNDFIVVELVKGYL 962
 QY 304 EFRYELGSLAVLR--SAPLALGRWHRVSAER-----LNKDGSLRVNGRPRVLS 352
 Db 963 HVYFDLGNANLIKGSNKPLNDNQHNVMISRDTSNHLTKVTKITQITAG----- 1016
 QY 353 SPKSGOGLNHLTLVLLGGV--EPSVPLSPATNMSAHRFGCVGEVYVNGKRLDLYSFLGS 410
 Db 1017 ---ARNLDLKSPLDLYIGVAKETYSKPLVHAKEGFGCLASVDLNGRLPDL----- 1065
 QY 411 QGIGQCYDSSPCERQPCQHGATCMPAGEYFQCLCRDGFKGLDCEHENPCQLREPCPLHG 470
 Db 1066 ---ISDALFCN-----CGQIE-----RG--CEGPTTTCQ--EDSCSNQ 1095
 QY 471 GTC----QGTRCLC-LPGFSGPRCQO-----GSGHGIAESDWHLEGGNDAPQGYGA 518
 Db 1096 GVCLOQWDSICDCSMTSFGPLCNDPDTGTYTFSKGGQITVKK-----PNDRPPSTRA- 1149
 QY 519 YFHDGFLAFPGHVRFSRSLPEVETIELE---VRTSTASGLLLMQGVEVGBAGQKDFIS 575
 Db 1150 ---DRLAIGF-----STVQKEAVLVRVDSSSGL-----GDYLE 1179
 QY 576 IGLQDGHVFRYQLGSGEARLVSEDP-INDGEWHRVYVTRALREGRGSIQVDG---BELVS 630
 Db 1180 LHHQKIGKVFNGVTDDIAIESNAIINDGKYHVVRFRSGGNATLQVDSWPVIERYP 1239
 QY 631 GRSPGNVAV-----NAKGSVYIGG----- 650
 Db 1240 GNDNERLAIARQIPYRLGRVVDWMLLDKGRQLTFNSQATIIIGKKEQGFQGLSG 1299

RESULT 10
 A40228
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Dec-2002
 R;Ushkaryov, Y.A.; Petrenko, A.G.; Geppert, M.; Suedhof, T.C.
 Science 257, 50-56, 1992
 A;Title: Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin recep
 A;Reference number: A40228; MUID:92320296; PMID:1621094
 A;Accession: A40228
 A;Status: preliminary; nucleic acid sequence not shown
 A;Molecule type: mRNA

651APDVATLGGFRSSGIGTCVKNLV 674
 1300 LYNGLKVLAENDANIALVGNVRLVGEVSPMTESTATAMQSEMSTSIEMETITTLA 1359

675 LHSARFGAPPQPLDLQHRQAQAGANTRFCPS 705
 1360 TSTARRGKPTKPERISQHTDDILVASAECPS 1390

RESULT 11
 T42626
 secreted leucine-rich repeat-containing protein SLI12 - mouse (fragment)
 N:Alternate names: neurogenic extracellular slit protein
 C:Species: Mus musculus (house mouse)
 C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 16-Aug-2002
 C:Accession: T42626
 R:Holmes, G.P.; Negus, K.; Burridge, L.; Raman, S.; Algar, E.; Yamada, T.; Little, M.H.
 Mech. Dev. 79, 57-72, 1998
 A:Article: Distinct but overlapping expression patterns of two vertebrate slit homologs in
 A:Reference number: Z22177; MUID:99279238; PMID:10349621
 A:Accession: T42626
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Gene: Slit2
 A:Residues: 1-1025 <HOL>
 A:Cross-references: EMBL:AF074960; NID:94151258; PID:94151259; PIDN:AAD04345.1
 C:Genetics:
 C:Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein

Query Match 10.1%; Score 387.5; DB 2; Length 1025;
 Best Local Similarity 29.7%; Pred. No. 4e-17;
 Matches 108; Conservative 55; Mismatches 144; Indels 57; Gaps 18;

Qy 158 HC-----PFCRRPCONGQCHDSSESYVCPAGFTGSRCEHSQAL-----HCHPEAC 207
 Db HCDIDFDDCODNCKNGAHCCTDA-VNGYTCVCEGYSGLFCFSPMVLPRTPSPCDNFD 626

Qy 208 GPDATC---VNRPDGRVTCRCHLGRSLRCEGVTVTTPSLGAGSYLALPAL-----T 259
 Db 627 QNGAQLIRNEP-----IOCLPEYLGEKELVSV---NFVKGESYLIQIPSAKVRPQT 678

Qy 260 NTHHELRLDVEFKPLAPDGVLLFSGKSGFVDFVSLAMVGGHLEFRYELGSGLA-VLRS 318
 Db 679 NITLQIATDE-----SGILLYKGDK-----DHIAVELYGRVRSYDTGSHPASAIYS 727

Qy 319 AEPLALGHRVSAERLNKOGSLRVNGRVPVLRSSPKSQGLNHLTLILYGGVE-----P 373
 Db 728 VETINDGNFHLVLELLTLDSSLSLSVDGSGPKVIINLSKQTLNPFSLYVGGMPGKNVA 787

Qy 374 SVPLSPATNMSAHRGCVGEVSNKRLDLYSPGSGGICQCVDSRCPCEHGCATC 433
 Db 788 SLRQAPGN-GTSPHGCIRNLYINSELQDFRKMVMQGIPLGC---EPCHKKYCAHG-MC 842

Qy 434 MPAGEYFQCLCRDGFKGDICEHEEN-PCQLRPELHGTCQ-----GTRCLCLPGFSGP 487
 Db 843 QPSSQSGFTCEEGWGLPCDQRTDPC-LGNKCVH-GTCLPINFAPSYCKCLEHGGV 900

Qy 488 RCOQ 491
 Db 901 LCDE 904

RESULT 12
 T13953
 MEGF5 protein - rat
 N:Alternate names: slit protein homolog
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 16-Aug-2002
 C:Accession: T13953
 R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
 Genomics 51, 27-34, 1998
 A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
 A:Reference number: Z14126; MUID:98360089; PMID:9693030

A:Accession: T13953
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1523 <NAK>
 A:Cross-references: EMBL:AB011531; NID:93449291; PIDN:BA032461.1; PID:93449292
 C:Genetics:
 A:Gene: MEGF5
 C:Superfamily: fruit fly slit protein; EGF homology; leucine-rich alpha-2-glycoprotein

Query Match 10.1%; Score 387; DB 2; Length 1523;
 Best Local Similarity 28.6%; Pred. No. 6.7e-17;
 Matches 113; Conservative 52; Mismatches 148; Indels 82; Gaps 18;

Qy 159 CPT-----CRDRPCONGQCHDSSESYVCPAGFTGSRCEHS-----QALLHCHPEACG 208
 Db 1071 CETDNDDCVAHKRGAQCVDV-VNGYTCVCEGYSGLFCFHPMVLQTSPCDQYEQ 1129

Qy 209 PDATCV---NRPDGRVTCRCHLGRSLRCEGVTVTTPSLGAGSYLALPALTNHHEL 265
 Db 1130 NGAQCIVVQEB-----TCRCPFGPAGPRCEKLTIV---NFVKGDSYVEL-ASAKVRPQA 1180

Qy 266 RLDVFKPLAPDGVLLFSGKSGFVDFVSLAMVGGHLEFRYE-LGSLAVLRSAPLAL 324
 Db 1181 NISLQVATDKONGIILLYKGD-----NDPLALELYQGHVLYVDSLSSEPTTVYSVETVND 1235

Qy 325 GRHVRVSAERLNKOGSLRVNGRVPVLRSSPKSQGLNHLTLILYGGVEVPSVPLSP-----A 380
 Db 1236 GQFHSVELVMLNQTLMVVDKGAPEKSLGKLOKQPAVGINSPYLLGGIPTSTGLSALROGA 1295

Qy 381 TNMSAHRGCVGEVSNKRLDLYSPGSGGIG----- 414
 Db 1296 DRPLCGFRGCIHEVRINNELQD--FKALPPQSLGVSPOCKSCTVCRHGLCRSVEKDSVVC 1353

Qy 415 QCYD-----SSPCERQPCOHGATCMPAGEYFQCLCRDGFKGDICEHEE---NP 460
 Db 1354 ECHPWGTGPLCDOEAQDFLGHSCSHG-TCVATGN-SYVCKAEGYGLPCDQKNSANA 1411

Qy 461 COLRPELHGTCQGTR---CLCLPGFSGPRCOQ 491
 Db 1412 CSAFK-CHHGQCHISDRGEPYCLCQPFSGNHCEQ 1445

RESULT 13
 IJFFTM
 cadherin-related tumor suppressor precursor - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Feb-1997
 C:Accession: A41087; B41087
 R:Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Biesemann, H.; Bryant, P.J.; Goodman, C.S.
 Cell 67, 853-868, 1991
 A:Title: The fat tumor suppressor gene in Drosophila encodes a novel member of the cadhe:
 A:Reference number: A41087; MUID:92069752; PMID:1959133
 A:Accession: A41087
 A:Molecule type: mRNA
 A:Residues: 143-485;1279-5147 <MAH>
 A:Cross-references: GB:M80537
 A:Accession: B41087
 A:Molecule type: DNA
 A:Residues: 1-142;487-1278 <MA2>
 A:Cross-references: GB:M80537
 A>Note: 1229-Gly and 1233-Ser were also found
 C:Genetics:
 A:Gene: fat
 A:Cross-references: FlyBase:FBgn001075
 C:Superfamily: cadherin-related tumor suppressor; cadherin repeat homology; EGF homology
 C:Keywords: calcium binding; cell adhesion; duplication; transmembrane protein
 F:1-35/Domain: signal sequence #status predicted <SIG>
 F:36-5147/Product: cadherin-related tumor suppressor #status predicted <MAT>
 F:36-4583/Domain: extracellular #status predicted <EXT>
 F:51-156/Domain: cadherin repeat homology <CR1>
 F:159-270/Domain: cadherin repeat homology <CR2>
 F:271-382/Domain: cadherin repeat homology <CR3>
 F:390-494/Domain: cadherin repeat homology <CR4>

F:497-599/Domain: cadherin repeat homology <CR5>
 F:602-708/Domain: cadherin repeat homology <CR6>
 F:718-822/Domain: cadherin repeat homology <CR7>
 F:831-942/Domain: cadherin repeat homology <CR8>
 F:948-1049/Domain: cadherin repeat homology <CR9>
 F:1052-1153/Domain: cadherin repeat homology <CR10>
 F:1156-1278/Domain: cadherin repeat homology <CR11>
 F:1281-1384/Domain: cadherin repeat homology <CR12>
 F:1387-1489/Domain: cadherin repeat homology <CR13>
 F:1492-1601/Domain: cadherin repeat homology <CR14>
 F:1607-1713/Domain: cadherin repeat homology <CR15>
 F:1717-1823/Domain: cadherin repeat homology <CR16>
 F:1826-1922/Domain: cadherin repeat homology <CR17>
 F:1925-2027/Domain: cadherin repeat homology <CR18>
 F:2028-2167/Domain: cadherin repeat homology <CR19>
 F:2169-2278/Domain: cadherin repeat homology <CR20>
 F:2281-2384/Domain: cadherin repeat homology <CR21>
 F:2387-2491/Domain: cadherin repeat homology <CR22>
 F:2494-2596/Domain: cadherin repeat homology <CR23>
 F:2599-2703/Domain: cadherin repeat homology <CR24>
 F:2707-2810/Domain: cadherin repeat homology <CR25>
 F:2813-2913/Domain: cadherin repeat homology <CR26>
 F:2915-3013/Domain: cadherin repeat homology <CR27>
 F:3014-3124/Domain: cadherin repeat homology <CR28>
 F:3127-3229/Domain: cadherin repeat homology <CR29>
 F:3232-3334/Domain: cadherin repeat homology <CR30>
 F:3337-3439/Domain: cadherin repeat homology <CR31>
 F:3442-3545/Domain: cadherin repeat homology <CR32>
 F:3548-3651/Domain: cadherin repeat homology <CR33>
 F:3654-3756/Domain: cadherin repeat homology <CR34>
 F:3954-4010/Domain: EGF homology <EG1>
 F:4017-4048/Domain: EGF homology <EG2>
 F:4056-4089/Domain: EGF homology <EG3>
 F:4096-4127/Domain: EGF homology <EG4>
 F:4584-4609/Domain: transmembrane #status predicted <TM>
 F:4610-5147/Domain: intracellular #status predicted <INT>

Query Match 10.0%; Score 383; DB 1; Length 5147;
 Best Local Similarity 22.9%; Pred. No. 4.7e-16;
 Matches 185; Conservative 97; Mismatches 265; Indels 260; Gaps 37;
 QY 4 ITFRPDSADGMLLYNGKRVKVPSPNL-----ANRQPDFISFGLVGGPFRERFD 52
 DB IKSRMSNEDEVLYLLEGGNSNTLQMLMAVRLAKTSYQPKILRLRKRKSAF--- 3939
 QY 53 AGSGMATIRHPTPLALGHFTVTLRLSTQGLIYV-----DLAPVNGTSGKFKGLD 105
 DB -----SELLQKEVIVGVEPCSEPDVCEVNGVCSATMRLLD 3974
 QY 106 LNE-----ELYLGG---YPPYGAIPKAGLSGGFIG--CVR-----BLRI 139
 DB 3975 AHSFVIQDSPALVSGPRVVDY-----SCQTSFGSGQCRRRQDCLPNCPSOVQCR 4030
 QY 140 QGEEIVFHDNLTAHGISHC-----PTCRDRPCQGGQCHDS-ESSSYVVCVCPAGFTGSR 193
 DB 4031 LQSD--FQCMCFANRDKRCKERSDVKSPKCRNGSSCQSPDGSYFCLCRPFRGNQ 4088
 QY 194 CEHSQALHCHPEAGPDATCVN-RPDGRTYCRHLRSGURCEBGEVVTTPPSLSGAGSY 252
 DB 4089 CE-SVSDSCRPNPLHGLGLCVSLKP---GYKNCNTPGRYRHCERFSYGFQPL-----SY 4139
 QY 253 LALPALNTHHELRLDVEPKLAPDGVLLFS-GGKSGPVEDVFSVLMVGGHLEFRYELG- 310
 DB 4140 MTFPALDVTND--ISIVPATTKPNSLLLYNMGSGGSRDFLAIELV--HRAVFFSSGG 4195
 QY 311 --SGLAVLRSAPELALGRVHRVSAERLNKDGSLRV-----NGRPVLRSSPKSQ----- 358
 DB 4196 ARTAISTVIAGRNLADGGWHKVTATRNGRVMLSVAKCADSGDVCTECLPGDSSCYADEV 4255
 QY 359 ---GLNHTL-LYLGGVPSVPL--SPATNSAHRFCVGEVSNVNGKRLDLYTSFGLSQ 411
 DB 4256 GPVGLTNFNKQPMIGGLSADPILRFGQVHSDLLVCLHSVHIGGRALNLS-SPLQQK 4314

QY 412 GIQCVDSSPCRPQCGHATCMPAGEYEFQCLCRDGFKGLDCEHENPCQIRBPCLIHG 471
 DB 4315 GI-----LAGCNRQACQ-----PALAAE 4332
 QY 472 TCGTRCLCPGFSRCPQCGSHGSIABSDWHLEGGGND--PCQYGAIFHDDGFLAAPP 529
 DB 4333 RCGFAGQCIDRWSSSLCCQCG-----HLQSPDCSDSLEPITLG---EGAFVFR 4379
 QY 530 -GHVFSR-----SLPEVPEFIELEVRTSTASG 555
 DB 4380 ISEIYRMLLDNLYNSKAWLNDQMRERRAVSNFSTASQIYEAPKMLSLFRYKQGG 4439
 QY 556 LLLMQGVEVGEAGQKDFISLGLQDGHVFRYQLGSGEARLVSE-DPINDGEWHRVIALR 614
 DB 4440 QILDY-----AATQWFTLSLRGRVLYYSKQHLTINMTVQETSLNDCKWHNVSLFS 4492
 QY 615 EGRRSIQVDGSELVSGRSPGNVAVNAKGSVYIIGADPVATL-----TGRGF 662
 DB 4493 EGRSLRLIYDGRQ-----VGDELDIAGVHDFLDPYLLTILNVGG-- 4530
 QY 663 SSGITGCYKLVLSHSGARPGAPPPPLD 689
 DB 4531 -EAFVGLANVTNNEL-----QPLN 4550

RESULT 14
 A48216
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Dec-2002
 C/Accession: A48216; B48216
 R/Ushkaryov, Y.A.; Suedhof, T.C.
 Proc. Natl. Acad. Sci. U.S.A. 90, 6410-6414, 1993
 A/Title: Neurexin IIIalpha: extensive alternative splicing generates membrane-bound and
 A/Reference number: A48216; MUID:93342001; PMID:8341647
 A/Accession: A48216
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1438 <USH>
 A/Cross-references: GB:L14851
 A/Accession: B48216
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1368,1372-1438 <US2>
 A/Cross-references: GB:L14851
 C/Genetics:
 A/Introns: 1372/1
 A/Supersfamily: neurexin; EGF homology
 C/Keywords: alternative splicing; brain; cell surface component; duplication; extracellu
 F;1-27/Domain: signal sequence #status predicted <SIG>
 F;202-234/Domain: EGF homology <EGF>
 F;651-683/Domain: EGF homology <EGF1>

Query Match 10.0%; Score 382; DB 2; Length 1438;
 Best Local Similarity 19.8%; Pred. No. 1.3e-16;
 Matches 175; Conservative 138; Mismatches 273; Indels 296; Gaps 38;
 QY 1 EIKITFRPDSADGMLLYNGKRVKVPSPNLANRQPDFISFGLVGGPFRERFDAGSG-MAT 59
 DB 286 EITLSEKTPWRGKLIILHTG-----KSDYVNLALKDGAVSLVNLGSGAFEA 332
 QY 60 IRHPTLALGHF-----HTVTLRLSTQGLIYVGLDAPVNGTSGKFKOGLDNEELYGG 114
 DB 333 IVEPVN---GKENDNAWHDVKVTRNLRLQVTSVVDGILLTTTGYTQEDYTMLGSDDFYVGG 389
 QY 115 YPDYGAIPKAGLSGGFIGVCRHLRSGURCEBGEVVTTPPSLSGAGSY 174
 DB 390 SPETADLPQSPVSNFMGCLKEVYKNDIRLESLRARIQ----- 430
 QY 175 DSSSYVVCVCPAGFTGSRCEHSQALHCHPEAGPDATCVNRPDGRYTCRHLRSGUR 234
 DB 431 DTRMKIY-----GEVVEK 443

Query Match 10.0%; Score 382; DB 2; Length 1471;
 Best Local Similarity 19.8%; Pred. No. 1.3e-16;
 Matches 175; Conservative 138; Mismatches 273; Indels 296; Gaps 38;

QY 1 EIKITPPDSADGMLLVNGQKRVPGSPTNLANRPPDIFGLVGGPEFRFDAGS-MAT 59
 DB 286 EITLSFTWQENGLLITG-----KADYVNLAKDGAIVSLINLGSAGFA 332

QY 60 IRIPTPLALGHP-----HTVTLRSITOGSLIVDILAPVNGTSGQKFOGLDNLNEELYLG 114
 DB 333 IVEFVN---GKFNDNAHWKVTNRNLRQVITISVDGILITTYQEDYTWLGSDDFFVVG 389

QY 115 YPDYGAIPKAGLSGFGVRELRIOEERIVFDHMLTAHIGSHCHCPTCRDRPCQNGQCH 174
 DB 390 SPETADLPFGSVNFWGCLKEVVYKNNDIRLELSRLRIG----- 430

QY 175 DSSSSIVCVCPAGTGSRCHEHSQALHCHPEACPDATCVNRPDGRGYTCRCHLGRSGLR 234
 DB 431 DTKMKIY----- 443

QY 235 CEEGVTVTPSLGAGSYLALPALTWTHHELRLDVEFKPLAPDQVLLFSGGKSPVE--- 291
 DB 444 CENVATLDPINFETPEAYISLPK-WNTKMGSIISDFRTEPENGILLFTHGK--PQERKD 500

QY 292 -----DFVSLAMVGGHLEFRYELGSLAVLRSAPLAL-GRWHRVSAERLNKDGSL 341
 DB 501 VRSQKTKVDFFAVELLDGNLYLLDMSGSTIKVKATQKANDGEWYHVDIQDRSGSTI 560

QY 342 RVNGGRVLRSSPKSGQLNHLTLVYLGVEPS-----VPLSPATNMSAH-FRGCVEVS 395
 DB 561 SVNSRRTPPTAS-GESEILDLEGMYLGGPEFRAGLILFTELWTLAMLYGYGCIIDL 619

QY 396 VNG--KRLDLTYFLGSGI-GQC--YDSSPCEROPCOHGATCMPAGEYFOCLCR-DGF 449
 DB 620 IDGRSKIRQLAEMQNAAGVKSSCRMSAKQDSYPCKNNAVKD-GWNRFCIDCTGTGY 678

QY 450 KGDLCHEEN-----PCQLREPC----- 467
 DB 679 WGRTCEREAISLYDGSMTYKMPVMVWHTAEADVSPRMSQRAYGILLVATTSRDSADTL 738

QY 468 ---LHGG-----TCQGRCLCLPFGSPPRCQSGHGIAESDWH----- 503
 DB 739 RLELDGGRVKLWNLDCIRINC--NSSKGPETLY-AGQKLNDEWHVTRVWRGKSLKL 794

QY 504 -----LEGSGNDAPGOYAYFH--DDGFL-----APPGHVFS----- 534
 DB 795 TVDDDDVAEGTMVGD---HTRLEFHNIETGIMTEKRYISVVVPSFIGHLQSLMFLYID 851

QY 535 -----RSLPEVPEITIE-----LEVRTSTASGL 556
 DB 852 LCKNGDIDYCELKARFGLRNIADPVTFKTKSSYLTLATLQAYTSMHLFPQFKTTSADGF 911

QY 557 LLWQGVVEGAGQKDFISLGLQDHLVFRYQLGSGEARLV--SEDPINDGEWHRVTLAR 614
 DB 912 ILF-----NSGDNDFIAVELVKGIHYVFDLGNPNVWIKGNSDRPLNDQHWNVITR 965

QY 615 E-GRRGSIQVDE---ELVSGRSPGNVAVNAKGSVYIGGAPDVATLTGGRFS----- 664
 DB 966 DNSNTHSLKVDYTKVTVQVINGAK-----NLDLKGDLYMAG-----LAQGMYSNLPKLV 1014

QY 665 ---GITGVKNLVLSARPGAPPPQPLDLOHRAQAGANTRPC 703
 DB 1015 SRDGFQGLASVDLN---GRLPLINDALHR--SQIDRGC 1050

Search completed: March 9, 2004, 17:22:34
 Job time : 28.1076 secs

QY 235 CEEGVTVTPSLGAGSYLALPALTWTHHELRLDVEFKPLAPDQVLLFSGGKSPVE--- 291
 DB 444 CENVATLDPINFETPEAYISLPK-WNTKMGSIISDFRTEPENGILLFTHGK--PQERKD 500

QY 292 -----DFVSLAMVGGHLEFRYELGSLAVLRSAPLAL-GRWHRVSAERLNKDGSL 341
 DB 501 VRSQKTKVDFFAVELLDGNLYLLDMSGSTIKVKATQKANDGEWYHVDIQDRSGSTI 560

QY 342 RVNGGRVLRSSPKSGQLNHLTLVYLGVEPS-----VPLSPATNMSAH-FRGCVEVS 395
 DB 561 SVNSRRTPPTAS-GESEILDLEGMYLGGPEFRAGLILFTELWTLAMLYGYGCIIDL 619

QY 396 VNG--KRLDLTYFLGSGI-GQC--YDSSPCEROPCOHGATCMPAGEYFOCLCR-DGF 449
 DB 620 IDGRSKIRQLAEMQNAAGVKSSCRMSAKQDSYPCKNNAVKD-GWNRFCIDCTGTGY 678

QY 450 KGDLCHEEN-----PCQLREPC----- 467
 DB 679 WGRTCEREAISLYDGSMTYKMPVMVWHTAEADVSPRMSQRAYGILLVATTSRDSADTL 738

QY 468 ---LHGG-----TCQGRCLCLPFGSPPRCQSGHGIAESDWH----- 503
 DB 739 RLELDGGRVKLWNLDCIRINC--NSSKGPETLY-AGQKLNDEWHVTRVWRGKSLKL 794

QY 504 -----LEGSGNDAPGOYAYFH--DDGFL-----APPGHVFS----- 534
 DB 795 TVDDDDVAEGTMVGD---HTRLEFHNIETGIMTEKRYISVVVPSFIGHLQSLMFLYID 851

QY 535 -----RSLPEVPEITIE-----LEVRTSTASGL 556
 DB 852 LCKNGDIDYCELKARFGLRNIADPVTFKTKSSYLTLATLQAYTSMHLFPQFKTTSADGF 911

QY 557 LLWQGVVEGAGQKDFISLGLQDHLVFRYQLGSGEARLV--SEDPINDGEWHRVTLAR 614
 DB 912 ILF-----NSGDNDFIAVELVKGIHYVFDLGNPNVWIKGNSDRPLNDQHWNVITR 965

QY 615 E-GRRGSIQVDE---ELVSGRSPGNVAVNAKGSVYIGGAPDVATLTGGRFS----- 664
 DB 966 DNSNTHSLKVDYTKVTVQVINGAK-----NLDLKGDLYMAG-----LAQGMYSNLPKLV 1014

QY 665 ---GITGVKNLVLSARPGAPPPQPLDLOHRAQAGANTRPC 703
 DB 1015 SRDGFQGLASVDLN---GRLPLINDALHR--SQIDRGC 1050

RESULT 15
 B48218
 neurexin III-alpha membrane-bound type 3 precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Dec-2002
 C:Accession: B48218; C48218
 R:Ushkaryov, Y.A.; Suedhof, T.C.
 Proc. Natl. Acad. Sci. U.S.A. 90, 6410-6414, 1993
 A:Title: Neurexin IIIalpha: extensive alternative splicing generates membrane-bound and soluble isoforms
 A:Reference number: A48216; MUID:93342001; PMID:8341647
 A:Accession: B48218
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1471 <USH>
 A:Cross-references: GB:L14851
 A:Accession: C48218
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1368,1372-1471 <US2>
 A:Cross-references: GB:L14851
 C:Genetics:
 A:Introns: 1372/1
 C:Superfamily: neurexin; EGF homology
 C:Keywords: alternative splicing; brain; cell surface component; duplication; receptor;
 F:1-27/Domain: signal sequence #status predicted <SIG>
 F:202-234/Domain: EGF homology <EGF>
 F:651-683/Domain: EGF homology <EGF1>

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:19:08 ; Search time 51.7431 Seconds
(without alignments)
2876.963 Million cell updates/sec

Title: US-10-006-011a-3
Perfect score: 3825
Sequence: 1 EIKITRPDSADGMLLYNGQ.....QPLDLQHRAGANTPCPS 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:
1: /cgm2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep.*
2: /cgm2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
3: /cgm2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep.*
4: /cgm2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep.*
5: /cgm2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
6: /cgm2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep.*
7: /cgm2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*
8: /cgm2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep.*
9: /cgm2_6/ptodata/1/pubppaa/US09A_PUBCOMB.pep.*
10: /cgm2_6/ptodata/1/pubppaa/US09B_PUBCOMB.pep.*
11: /cgm2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*
12: /cgm2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep.*
13: /cgm2_6/ptodata/1/pubppaa/US10A_PUBCOMB.pep.*
14: /cgm2_6/ptodata/1/pubppaa/US10B_PUBCOMB.pep.*
15: /cgm2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep.*
16: /cgm2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep.*
17: /cgm2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*
18: /cgm2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Table with columns: Result No., Score, Query Match, Length, DB ID, Description. Contains 18 rows of search results.

Table with columns: 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45. Contains alignment data and accession numbers.

ALIGNMENTS

RESULT 1
US-10-094-886-180
Sequence 180, Application US/10094886
Publication No. US20040002120A1
GENERAL INFORMATION:
APPLICANT: Tchernev, Ramesh
APPLICANT: Liu, Xiaohong
APPLICANT: Spytek, Kimberly A.
APPLICANT: Burgess, Catherine
APPLICANT: Vernet, Corine A.
APPLICANT: Li, Li
APPLICANT: Gorman, Linda
APPLICANT: Malyankar, Uriel M.
APPLICANT: Boldog, Ferenc
APPLICANT: Guo, Xiaojia
APPLICANT: Shenoy, Suresh
APPLICANT: Padigar, Muralidhara
APPLICANT: Taupier, Raymond J., Jr.
APPLICANT: Miller, Charles
APPLICANT: Casman, Stacie
APPLICANT: Pena, Carol
APPLICANT: Gargolli, Esha
APPLICANT: Gusev, Vladimir
APPLICANT: Smithson, Glennnda
APPLICANT: Zerhusen, Bryan
APPLICANT: Gerlach, Valerie
APPLICANT: Pochart, Pascal
APPLICANT: Fernandes, Elma
APPLICANT: Shimkets, Richard
APPLICANT: Rastelli, Luca
APPLICANT: Spaderna, Steven
APPLICANT: Larochele, William
APPLICANT: Zhong, Mei

TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
FILE REFERENCE: 21402-290 B
CURRENT APPLICATION NUMBER: US/10/094,886

; CURRENT FILING DATE: 2002-03-07
 ; PRIOR APPLICATION NUMBER: 60/274,322
 ; PRIOR FILING DATE: 2001-03-08
 ; PRIOR APPLICATION NUMBER: 60/313,182
 ; PRIOR FILING DATE: 2001-08-17
 ; PRIOR APPLICATION NUMBER: 60/288,052
 ; PRIOR FILING DATE: 2001-05-02
 ; PRIOR APPLICATION NUMBER: 60/318,510
 ; PRIOR FILING DATE: 2001-09-10
 ; PRIOR APPLICATION NUMBER: 60/274,281
 ; PRIOR FILING DATE: 2001-03-08
 ; PRIOR APPLICATION NUMBER: 60/314,018
 ; PRIOR FILING DATE: 2001-08-21
 ; PRIOR APPLICATION NUMBER: 60/274,194
 ; PRIOR FILING DATE: 2001-03-08
 ; PRIOR APPLICATION NUMBER: 60/274,849
 ; PRIOR FILING DATE: 2001-03-09
 ; PRIOR APPLICATION NUMBER: 60/296,693
 ; PRIOR FILING DATE: 2001-06-07
 ; PRIOR APPLICATION NUMBER: 60/313,626
 ; PRIOR FILING DATE: 2001-08-21
 ; Remaining Prior Application data removed - See file Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 298
 ; SOFTWARE: PatentIn 2.1
 ; SEQ ID NO 180
 ; LENGTH: 1931
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-094-886-180

QY 556 LLIWQGVVEGEAGQKDFISIGLQDHLVFRYQLGSGEARLVSEDPINDEWHRVTALRE 615
 Db 1791 LVLWS---GKATERADYVALAIVDGHLQLSYNLGSPVVLRSYVTVNTRMLRVVAHRE 1846
 QY 616 GRGSIQVDEBELVSRSPGNVAVNAKGSVYIGGAPD--VATLTGGRFSSSITGCVKML 673
 Db 1847 QREGSLQVGNAPVTGSSPLGATQDLDGALWLGPELFPVGPALPKAYGTGVGCLRDV 1906
 QY 674 VLHSARFGAPPQPLDQHRQAQAGANTRPCPS 705
 Db 1907 VVGR-----HPLLELLEDAVTKELRCPCT 1930
 RESULT 2
 US-10-016-283-34
 ; Sequence 34, Application US/10016283
 ; Publication No. US20020164702A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela et al., David M.
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
 ; FILE REFERENCE: REG195-B-PCT-US
 ; CURRENT APPLICATION NUMBER: US/10/016,283
 ; PRIOR FILING DATE: 2001-11-30
 ; PRIOR APPLICATION NUMBER: US/09/077,955A
 ; PRIOR FILING DATE: 1998-09-10
 ; PRIOR APPLICATION NUMBER: PCT/US96/20696
 ; PRIOR FILING DATE: 1996-12-13
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 34
 ; LENGTH: 1940
 ; TYPE: PRT
 ; ORGANISM: Rattus sp.
 US-10-016-283-34

Query Match 23.6%; Score 901.5; DB 15; Length 1931;
 Best Local Similarity 32.2%; Pred. No. 1.7e-64;
 Matches 242; Conservative 93; Mismatches 266; Indels 151; Gaps 22;
 QY 2 IKITRPDSADGMLLYNQKRVGSPPTLANRQDFISFGLVGRPEFRFDAGSGMATIR 61
 Db 1282 LALEFRALTEGELLYNGNA-----RGKDFLALALDDGRVQRFDTGSGPAVLT 1330
 QY 62 HPTPLALGHFHTVLLRSLTQGLSIVGDLAPVNGTSQKFGQLDNLNEELYLGGYD--Y 118
 Db 1331 SAVFVPCQWRHLELSRHRRTLSVDGETVLSGESPGTDGLNLDLDFVGGVPEQAA 1390
 QY 119 GAIPKAGLSSGFIGVRELRIGQBEIVFHDNLN-----TAHGISHC--PTCRDRPCQNG 170
 Db 1391 VALERTFVAGLRGICRILLDNNQRLL---ELGIFGAATRSGVGECDHPCLNPNCHGG 1447
 QY 171 QCHDSESSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRDPDGRGYTCRCHLGR 230
 Db 1448 APCQMLEAGRPHCCPPGRVGPCTCADEKS--PCQNPCHGAAPCRVLPFG--GAQCEPLGR 1505
 QY 231 SGLRCEEGVYTTTSLSGAGSYLALPALTNTHHEL-----LDVEFKPLAP 276
 Db 1506 EGTEFCQ-----TASGQDGGSPFLA--DFNGFSLHLRGLHTFARDLGEKMALEVYVFLARGP 1559
 QY 277 DGVLLFSGKSGPVEDFVSLAMVGGHLEFRYELGSLAVLRSABPALGRWHRVSAERLN 336
 Db 1560 SGLLYNGQKTDGKGDVSLALRDRLEFRYDLGKGAAVRSRFPVLTGAWTRVLERNG 1619
 QY 337 KDGSLRVNGRVPVLRSPGKSGQLNHLTLVYLGVEBVSFLSPATNMSAHRFCVGEVSV 396
 Db 1620 RKGALRVGDGPRVLGESPVHTVNLKLEPLVYVGGAPDFSKLARAASVSGFDGAIQLVSL 1679
 QY 397 NGKEL-----DLTYSPFGSGGICQCYDSSPCER---QPCQHGATCWPAGEYFQC 443
 Db 1680 GGRQLLPEHVLRQVDVT--SFAG-----HPCTRASHGPHCLNGASCVP----- 1720
 QY 444 LCRDGFKGLDCEHENPQLREPLCHGTCQGTTRCLCLPGFSGPRCQOQSGHGIAESDMH 503
 Db 1721 -----REAY-----VCLCPGGFSGPHCEKGL----- 1742
 QY 504 LEGSGNDAPQYAYFDDGFLAFPGHVRSRSLPEVPEP-----IELVRTSTAG 555
 Db 1743 VEKSAG-----DVDTLAFDGRTFVYLVNVAVTESEKALQSNHNFELSLRTEATQG 1790

Query Match 22.5%; Score 862.5; DB 13; Length 1940;
 Best Local Similarity 32.0%; Pred. No. 2.8e-61;
 Matches 239; Conservative 92; Mismatches 280; Indels 135; Gaps 22;
 QY 2 IKITRPDSADGMLLYNQKRVGSPPTLANRQDFISFGLVGRPEFRFDAGSGMATIR 61
 Db 1287 LALEFRALTEGELLYNGNA-----RGKDFLALALDDGRVQRFDTGSGPAVLT 1335
 QY 62 HPTPLALGHFHTVLLRSLTQGLSIVGDLAPVNGTSQKFGQLDNLNEELYLGGYDYG-- 119
 Db 1336 SLVPEPCQWRHLELSRHRRTLSVDGETVLSGESPGTDGLNLDLNLVYGGIPEEQVA 1395
 QY 120 AIPKAGLSSGFIGVRELRIGQBEIVFHDNLNFTA--HGI-SHC--PTCRDRPCQNGQC 173
 Db 1396 MVLDRTSVGVGLKGCIRMLDINNQQLELSDMQRAAVQSSGVGECGDHPCLNPNCHGGALC 1455
 QY 174 HDSESSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRDPDGRGYTCRCHLGRSGL 233
 Db 1456 QALEAGMFLCQCPGRVGPCTCADEKS--PCQNPCHGAAPCRVLSGG--GAKCECPLGRSGT 1513
 QY 234 RCEEGVYTTTSLSGAG-----SYLALPALTNTHHEL-----RLDVEFKPLAPDG 278
 Db 1514 FCQ---TV-----LETAGSRPFLADPNFGFSYLELGLHTFERDLGKMALEVMVFLARGPSSG 1566
 QY 279 VLLFSGKSGPVEDFVSLAMVGGHLEFRYELGSLAVLRSABPALGRWHRVSAERLNKD 338
 Db 1567 LLLYNGQKTDGKGDVSLALRDRLEFRYDLGKGAAVRSRKEPALTGTWVTVFLERNGR 1626
 QY 339 GSLRVNGRVPVLRSPGKSGQLNHLTLVYLGVEBVSFLSPATNMSAHRFCVGEVSV 393
 Db 1627 GALQVGDGPRVLGESPVHTVNLKLEPLVYVGGAPDFSKLARGAASVSGFSGVQL 1685
 QY 394 VSVNGKRLDLYTFLGSGGICQCYDSSPCERQPCQHGATCWPAGEYFQCLCRDGFKGLD 453
 Db 1686 VSLRSHQL-----LTQEHVLRVDVDSFADHPCTQA----- 1716
 QY 454 CEHEENPCQLREPLCHGTC---QGT--RCLCLPGFSGPRCQOQSGHGIAESDMHLEGGG 509

Db 1717 -----LGNPCLNGSCVREATYECLECGGSLHCKGL-----VEKSVG 1757
 QY 510 NDAPGOYGFYHDDGFLAPFGHVFVRSLEPEVET-----IELEVRSTASGLLLWOG 561
 Db 1758 -----DLETAFDGRTYIEYLNAVISESEKALQSNHFELSRLATEAQLVW-- 1803
 QY 562 VEYGEAGQKDFISLQDGHVFRYQLGSGEARLVSEDPINDGWHHRVTVTLRGRGSI 621
 Db 1804 --IGKAERADYALAIVDGHLQLSDVLSQPVLRSYKVNTRWLRIRAHREHREGSL 1861
 QY 622 QVDGELVSGRPGPNVAVNAKSVYIGAP--DVATLTGGRFSSGITCVKKNLVLSHAR 679
 Db 1862 QVNEAPVTVSSPLGATQDLDGALWGLQKLPVGOALPKAYGTFVGCCLRDVVVGH-- 1919
 QY 680 PGAPPEPLDQHRAGAGANTRPCPS 705
 Db 1920 -----RQLHLEDAVTKBELRPCPT 1939

RESULT 3
 US-10-108-260A-4433
 ; Sequence 4433, Application US/10108260A
 ; Publication No. US20040005560A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HELIX RESEARCH INSTITUTE
 ; TITLE OF INVENTION: No. US20040005560A1e1 full length cdna
 ; FILE REFERENCE: H1-A0106
 ; CURRENT APPLICATION NUMBER: US/10/108,260A
 ; CURRENT FILING DATE: 2002-03-27
 ; NUMBER OF SEQ ID NOS: 5458
 ; SOFTWARE: Patent in Ver. 2.11
 ; SEQ ID NO 4433
 ; TYPE: PR1
 ; ORGANISM: Homo sapiens
 US-10-108-260A-4433

Query Match 22.0%; Score 840; DB 15; Length 775;
 Best Local Similarity 29.6%; Pred. No. 5.4e-60;
 Matches 214; Conservative 107; Mismatches 257; Indels 144; Gaps 20;
 QY 1 EIKITFRPDSADGMLLYNQKRVPGSPPTLANRQDDPISFGLVGRGPRFRFDAGSGMATI 60
 Db 176 QITLFRAEABDGLLYCGEEN-----HGRGDFMSLAIIRSLQFRNCGTGVAIL 226
 QY 61 RHPTPLALGHFHTVTLRSLTQGSLLVGDLPVNGTSQKFGQLDLNEELVLYGYPD-YG 119
 Db 227 VSETKIKLGGWHTVMLYRDGLNGLLQLNNGTPTVTSQSQYKSKITFRTPLYLGGAPSAW 286
 QY 120 AIPKAGLSSGFTGCVRELRIOGEEIVFHDLN-----LTAHGISHCPT--CRDRPCQNG 170
 Db 287 LVRATGTRNGFQGCQSLAVNGKRI---DMRPWPLGKALSADYGECCSSGICDEASCING 343
 QY 171 GOCHDSESSYVVCVPAGFTGSRCHSQAALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
 Db 344 GTCATAIKADSYICLCLPLGFR-----362
 QY 231 SGLRCEGVVTTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFSGK 283
 Db 363 XGRHCEDAFTLTIPOFRSRLSYAATPWLEPQHYLSFMEFEITFRPDSGGVLLYSYDT 422
 QY 284 GKSGPVEDFVSLAMVGHLEFRYELGSLAVLSRAEPLALGRWHRVSAERLNKDGSLRV 343
 Db 423 GSK-----DFLSINLAGHVEFRPDCSGTGLVRSDEPLTLGNWHELRSVSTAKNGILQV 477
 QY 344 NGRVLRSSPKSOGMLNHLILLYGGVEPSVPLSPATNMAHFRGCVGEVSVNGKRLDL 403
 Db 478 DKQKIVEGMAEGGFTQIKCNTDIFIGVFNVDVVKNSGVLKPPFSGTQKILLNDRTHV 537
 QY 404 TVSFVLGSGIGQCYDSSPCERPOCHGATCWPAGRYEFCQLCRDGFKGDLCHEBENFCQL 463
 Db 538 XHDF--TSGVNVNNAHFVCRAPCAHGGSCAPRKE-GYDCCDCLPLGFB-----582

QY 464 REPCLHGCTCGT--RCLCLPFGSRRCQOQSGHGIAESDWHLEGGSGNDAPGOYGAYFH 521
 Db 583 ----LH---CKKAIIEAIEIPQFIG-----RSYLTYN 608
 QY 522 DDGFLAPFGHVFVRSLEPEVETTELEVRTSTASGLLLWOGVEYGEAGQKDFISLQDGH 581
 Db 609 PDILKRVSG--SRS-----NVFMFKITAKDGLLWRG--DSPMRPNDFDLSLGLRDG 657
 QY 582 HLVFRYQLGSGEARLVSEDPINDGWHHRVTVTLRGRGSIQVDGEBELVSGRSPGNVAVN 641
 Db 658 ALVFSYNLGSGVASIMVNGSFDNRWHRVKAVRDSQSGKITVDDYGARTKSPGWRQLN 717
 QY 642 AKGSVITGAPDVATLTGGRFSSGITCVKKNLVLSHARFCAPPPLDQLQHRAGAGANTR 701
 Db 718 INGALVYGGMKEIALTHTNRQIMRGLVCCISHFTLST-----DYHISLVEDAVDGRKNIN 770
 QY 702 PC 703
 Db 771 TC 772

RESULT 4
 US-09-866-050A-503
 ; Sequence 503, Application US/09866050A
 ; Publication No. US20030040471A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Watson, James D.
 ; APPLICANT: Strachan, Lorna
 ; APPLICANT: Sleeman, Matthew
 ; APPLICANT: Onrust, Rene
 ; APPLICANT: Murison, James G.
 ; APPLICANT: Kumble, Krishanand D.
 ; TITLE OF INVENTION: Compositions Isolated From Skin Cells
 ; FILE REFERENCE: 11000.1011c4U
 ; CURRENT APPLICATION NUMBER: US/09/866,050A
 ; CURRENT FILING DATE: 2001-05-24
 ; NUMBER OF SEQ ID NOS: 725
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 503
 ; LENGTH: 819
 ; TYPE: PR1
 ; ORGANISM: Rat
 US-09-866-050A-503

Query Match 21.8%; Score 832; DB 10; Length 819;
 Best Local Similarity 29.3%; Pred. No. 2.6e-59;
 Matches 210; Conservative 109; Mismatches 264; Indels 134; Gaps 17;
 QY 1 EIKITFRPDSADGMLLYNQKRVPGSPPTLANRQDDPISFGLVGRGPRFRFDAGSGMATI 60
 Db 220 QITLFRAEABDGLLYCGEEN-----HGRGDFMSLAIIRSLQFRNCGTGVAIL 270
 QY 61 RHPTPLALGHFHTVTLRSLTQGSLLVGDLPVNGTSQKFGQLDLNEELVLYGYPD-YG 119
 Db 271 ISETKIKLGGWHTVMLYRDGLNGLLQLNNGTPTVTSQSQYKSKITFRTPLYLGGAPSAW 330
 QY 120 AIPKAGLSSGFTGCVRELRIOGEEIVFHDLN-----LTAHGISHCPT--CRDRPCQNG 170
 Db 331 LVRATGTRNGFQGCQSLAVNGKRI---DMRPWPLGKALSADYGECCSSGICDEASCING 387
 QY 171 GOCHDSESSYVVCVPAGFTGSRCHSQAALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
 Db 388 GTCATAIKADSYICLCLPLGFR-----409
 QY 231 SGLRCEGVVTTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFSGK 286
 Db 410 --HCEBDAFTLTIPOFRSRLSYAATPWLEPQHYLSFMEFEITFRPDSGGVLLYSYDT 466
 QY 287 SGPVEDFVSLAMVGHLEFRYELGSLAVLSRAEPLALGRWHRVSAERLNKDGSLRVNG 346
 Db 467 SS--KDFLSINLAGHVEFRPDCSGTGLVRSDEPLTLGNWHELRSVSTAKNGILQV 524

```

QY 347 RVLRLSSPKSGLWHLTLVLLGGVPSVPLSPATNMSAHRFCVCVGEVSNKRDLDTYS 406
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 525 KVEGMABGGFTQKCNFTDPIGGVFNVDVYKNSGILHPFSGIQKIILNDRTHIVRHD 584
QY 407 FLGSGICQCCSDSPCEQOHCATMPAGEYFQCLCRDGFKDKGDLCEHEENPCQLREP 456
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 585 P--TSVVNVAHAHCYCAPCAHGGSCFRKE-GYECDCPLGFSGLNCOKAI----- 633
QY 467 CLHGTCCGTCLCLPFSGPRCQGGHGTABSDWHLEGGGNDAPQCYGAFHDDGFL 526
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 634 -----TEAIFQPIGRSYLTYDNPNLK---RVSGSRN-----AFM 668
QY 527 APFGHVFSRSLPEVPEITTELEVRTSTASGLLWQVGEVGEAGQKDFISLGLDGHVFR 586
DB 669 RP-----KITAKDGLLLWRG--DSPMRFNSDFISLGLRDRGALVFS 706
QY 587 YQLSGEARLVSEDPINDGEHWHRVTLRREGRRGSIQVDBELVSGRSPGNVAVNAKGSV 646
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 707 YNLGSGVASIMVNGSFSDGRWHRVYKAVRDSQSGKITVDDYGAARTKSGSNMQRQLNMGAL 766
QY 647 YIGGAPDVATLTGRFSSGITGCVKNLVLSHARFGAPPQLDLQHRQAQAGANTRPC 703
DB 767 YVGGMKEALTHNRQYRWGLVCCISHFLTST-----DYHISLVEDAVDKNIINTC 816

RESULT 5
US-10-016-283-36
; Sequence 36, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCY-US
; CURRENT APPLICATION NUMBER: US/10/016,283
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US/09/077,955A
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: PCY/US96/20696
; PRIOR FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 36
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-283-36

Query Match 16.7%; Score 640.5; DB 13; Length 492;
Best Local Similarity 33.0%; Pred. No. 6.6e-44;
Matches 183; Conservative 59; Mismatches 173; Indels 139; Gaps 18;

QY 202 CHPEACGDPATCVNRPDGRGYTCCHLGRSLRCEEVTVTTSPSLSGAGSYLALPALNVT 261
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 262 FHELR-----LDVFEKPLAPDGVLLFLSFGKSGPVDFVSLAMVGHLHLEFFRY 307
DB 80 HLELRGLHTFARDLGERMALEVFVLARGPSGLLNYNKTGKDGDFVSLALRBRLEFFRY 139
QY 308 ELGSLAVLRSAEPLALGRWHRVSAERLNDKSLRUVNGRPPVLRSSPGKQG-----LNL 362
DB 140 DLGKAAVTRRREFTLGAATRVSLERNRGAALRVGDPVRLGESP-KSRKVPHTVLANL 198
QY 363 HTLLYLGGEVPSVPLSPATNMSAHRFCVCVGEVSNKRL-----DLTYSELGSGQ 412
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 199 KEPLYVGGAPPFSKLARAAVSSGFDGAIQLVSLGGKQLITPEHLVROVDVT-SFAG--- 254
QY 413 IGQCYSDSPBER---PQCOHGATCMPAGEYFQCLCRDGFKDLCEHEENPCQLRPFCLH 469
DB 255 -----HPCYTRASGHPCLNGASCVP-----REAAV- 278
QY 470 GGTCCGTRCLCLPFSGPRCQGGSHGIABSDWHLEGGGNDAPQCYGAFHDDGFLAP 529

```

```

DB 279 -----VLCDFGSGPHCEKGL-----VEKSAG-----DVDTLAPD 309
QY 530 GHVFSRSLPEVPEIT-----LELVRTSTASGLLWQVGEVGEAGQKDF 573
DB 310 GRTFVEYLNAVTESELANEIPVEKALQSNHFELSRLTEATQGLVLMW---GKATERADY 365
QY 574 ISLGLQDGHVFRYQLSGEARLVSEDPINDGEHWHRVTLRREGRRGSIQVDBELVSGRS 633
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 366 VALAIVDGHLOJSYNLGSOQVVVLRSTVFNTRMLRVVAHREQRGGSLQVGNRAPVTGSS 425
QY 634 FGNVAVNAKGSVYICGGAPD-VATLTGGRFSSGITGCVKNLVLSHARFGAPPQLDLQ 691
DB 426 PLGATQDLDGALGLGGLPELVPGLPKAYGTGVPVGLRDRVYVGR-----HPLHL 477
QY 692 HRAQAGANTRPCPS 705
DB 478 EDAVTKPELRPCPT 491

RESULT 6
US-10-104-047-3058
; Sequence 3058, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1e1 full length cdna
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3058
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3058

Query Match 16.4%; Score 628; DB 15; Length 463;
Best Local Similarity 28.7%; Pred. No. 6.4e-43;
Matches 162; Conservative 80; Mismatches 198; Indels 124; Gaps 16;

QY 151 LTAHGISHCPT--CDDREFCQGGQCHDSSESSYVCVCPAGFTGSRCEHSQALHCHPEACG 208
DB 10 LSGADVGCSSGICDEASCIRHGTCTAIKADSYICLCPGF----- 50
QY 209 PDATCVNRPDGRGYTCRCHLGRSLRCEEVTVTTSPSL-SGAGSYLALPALNTHHEL-- 265
DB 51 -----KGRHCEADAFLLTIPQRESLRVSAATPWPELEPQHYLSF 88
QY 266 -RLDVEFKPLAPDGVLLFSS---GKSGPVDFVSLAMVGHLHLEFFRYELGSLAVLRSAEP 321
DB 89 MBEFTFRPDSGDGVLVSYDTSK-----DFLSINLAGHVEFRFCGSGTGLRSEDP 143
QY 322 LALGRWHRVSAERLNDKSLRUVNGRPPVLRSSPKSQGLNLHTLLYLGGVPSVPLSPAT 381
DB 144 LTLGNWHLRVSRAKNGIQLQVSKIVGVMAESGFTQIKCNTDIFIGGVFNVDVYKNS 203
QY 382 NMSAHRFCVCVGEVSNKRLDLYSFSLGSGIQCYDSSPCERQPCOHGATCMPAGEYEF 441
DB 204 GVLKFPFSGIKIIILNDRTHIVRHDFT--TSWVVENAAHPCVRAPCAHGGSCFRKE-GY 260
QY 442 QCLCRDGFKDLCEHEENPCQLRBPCLHGGTCCQT--RCLCLPGFSGPRCQGGSHGIAE 499
DB 261 DCCDFLGGEG-----LH-----CQKAIETAIPEIFIG----- 288
QY 500 SDWHLGSGGNDAPQCYGAFHDDGFLAPFGHVFSRSLPEVPEITTELEVRTSTASGLLW 559
DB 289 -----RSYLYTDNPDILKRVSG---SR-----NVFMFKTITAKDGLLW 325
QY 560 QGVVEVGEAGQKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEHWHRVTLRREGRRG 619

```

Db 326 RG--DSPMPNSDFISLGLDGLVSYNLSGVSIMVNGSFNDGRWHRVAVRDGQSG 383
 QY 620 SIQVDBELVSGSPGNVAVNAKAGSYVIGGAPDVATLTGGRFSSGITCVKXNVLHRSAR 679
 Db 384 KITVDDYGARTGKSPGMRNLNGALVYGMKEIALHTNRQYMRGLVGCISHFTLST-- 441
 QY 680 PGAPPPDLOHRAQAAGANTRPC 703
 Db 442 -----DYHISLVEDAVDGKNITC 460

RESULT 7
 ; US-09-764-853-541
 ; Sequence 541, Application US/09764853
 ; Patent No. US20020090672A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: P/J206
 ; CURRENT APPLICATION NUMBER: US/09/764,853
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 939
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 541
 ; LENGTH: 432
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-764-853-541

Query Match 15.8%; Score 605; DB 9; Length 432;
 Best Local Similarity 29.0%; Pred. No. 4.5e-41;
 Matches 157; Conservative 75; Mismatches 188; Indels 122; Gaps 15;

QY 171 GOCHDSESSYVVCVPAGTGRSCEHSQALHCHPEACGPDATCVNRPDGRGTYCCHLGR 230
 Db 1 GTCATKADSYICLCLPFG----- 19

QY 231 SGLRCEGVVTPPSL--SGAGSYLALPALNTWHEL---RLDVEFKPLADGVLLFS--- 283
 Db 20 KGRHCEDAFTLTIPOFRESLRSYAATPWLEPOHYLSFMEFIFTRPDSGDGVLVSYDT 79

QY 284 GKSGPVEDFVSLAMVGGHLEFRYELGSLAVLRSAPLALGRWHRVSAERLNKDGSLRV 343
 Db 80 GSK-----DFLSINLAGHVEFRFCGSGTVLRSDDPLTLGNWHELRSRTAKNGILQV 134

QY 344 NGRPVLRSSPGKSQGLNLHTLLYLGGVPSVPLSPATNNSAHFRGCVGEVSVNGKRLDL 403
 Db 135 DKQKIVEGMAEGGFTQIKCNTDIFIGVFNVDVKNKSGVLKPFSSIQKIILNDRTHV 194

QY 404 TYSFLGSGIGQCYSDBSPERPCQHGATCMPAGEYFQCLCRDGFKGDLCHEENPCQL 463
 Db 195 KHDF--TSGVNVNAAHPCVRAPCAHGGSCRPRKE--GYDCDCPLGFE----- 239

QY 464 REPCFHGGTCCGT--RCLCLPFGSPRCOQSGHGIAESDWELEGSGNDAPQYGFYH 521
 Db 240 ----LH--COKAIEAIEIPQFIG-----RSYLYTDN 265

QY 522 DDGFLAPGHVFRSLPEVETTELEVRTSTASGLLLMQGVEVGEAGQKDFISLGLQDG 581
 Db 266 PDILKRVSG---SRS-----NVFMRFKTKADGILLWRG--DSPMPNSDFISLGLRDG 314

QY 582 HLVPYQLSGEARLYSEDPINDGEWHRVTLRREGRRGSIQVDGEBELVSGRSPGNVAVN 641
 Db 315 ALVFSYMLSGSVASIMVNGSFNDGRWHRVAVRDGQSGKITVDDYGARTGKSPGMRLQN 374

QY 642 AKGSVYIGGAPDVATLTGGRFSSGITCVKXNVLHRSARFAGAPPQPLDLOHRAQAAGANTR 701
 Db 375 INGALYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DYHISLVEDAVDGKNIN 427

QY 702 PC 703
 Db 428 TC 429

RESULT 9
 ; US-09-764-881-102
 ; Sequence 102, Application US/09764881
 ; Publication No. US20030125246A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.

Db 428 TC 429

RESULT 8
 ; US-09-764-898-179
 ; Sequence 179, Application US/09764898
 ; Patent No. US20020090673A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: P/J201
 ; CURRENT APPLICATION NUMBER: US/09/764,898
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 311
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 179
 ; LENGTH: 432
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-764-898-179

Query Match 15.8%; Score 605; DB 9; Length 432;
 Best Local Similarity 29.0%; Pred. No. 4.5e-41;
 Matches 157; Conservative 75; Mismatches 188; Indels 122; Gaps 15;

QY 171 GOCHDSESSYVVCVPAGTGRSCEHSQALHCHPEACGPDATCVNRPDGRGTYCCHLGR 230
 Db 1 GTCATKADSYICLCLPFG----- 19

QY 231 SGLRCEGVVTPPSL--SGAGSYLALPALNTWHEL---RLDVEFKPLADGVLLFS--- 283
 Db 20 KGRHCEDAFTLTIPOFRESLRSYAATPWLEPOHYLSFMEFIFTRPDSGDGVLVSYDT 79

QY 284 GKSGPVEDFVSLAMVGGHLEFRYELGSLAVLRSAPLALGRWHRVSAERLNKDGSLRV 343
 Db 80 GSK-----DFLSINLAGHVEFRFCGSGTVLRSDDPLTLGNWHELRSRTAKNGILQV 134

QY 344 NGRPVLRSSPGKSQGLNLHTLLYLGGVPSVPLSPATNNSAHFRGCVGEVSVNGKRLDL 403
 Db 135 DKQKIVEGMAEGGFTQIKCNTDIFIGVFNVDVKNKSGVLKPFSSIQKIILNDRTHV 194

QY 404 TYSFLGSGIGQCYSDBSPERPCQHGATCMPAGEYFQCLCRDGFKGDLCHEENPCQL 463
 Db 195 KHDF--TSGVNVNAAHPCVRAPCAHGGSCRPRKE--GYDCDCPLGFE----- 239

QY 464 REPCFHGGTCCGT--RCLCLPFGSPRCOQSGHGIAESDWELEGSGNDAPQYGFYH 521
 Db 240 ----LH--COKAIEAIEIPQFIG-----RSYLYTDN 265

QY 522 DDGFLAPGHVFRSLPEVETTELEVRTSTASGLLLMQGVEVGEAGQKDFISLGLQDG 581
 Db 266 PDILKRVSG---SRS-----NVFMRFKTKADGILLWRG--DSPMPNSDFISLGLRDG 314

QY 582 HLVPYQLSGEARLYSEDPINDGEWHRVTLRREGRRGSIQVDGEBELVSGRSPGNVAVN 641
 Db 315 ALVFSYMLSGSVASIMVNGSFNDGRWHRVAVRDGQSGKITVDDYGARTGKSPGMRLQN 374

QY 642 AKGSVYIGGAPDVATLTGGRFSSGITCVKXNVLHRSARFAGAPPQPLDLOHRAQAAGANTR 701
 Db 375 INGALYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DYHISLVEDAVDGKNIN 427

QY 702 PC 703
 Db 428 TC 429

RESULT 9
 ; US-09-764-881-102
 ; Sequence 102, Application US/09764881
 ; Publication No. US20030125246A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PT207
 ; CURRENT APPLICATION NUMBER: US/09/764,881
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - refer to PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 192
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 102
 ; LENGTH: 432
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-764-881-102

Query Match 15.8%; Score 605; DB 10; Length 432;
 Best Local Similarity 29.0%; Pred. No. 4.5e-41;
 Matches 157; Conservative 75; Mismatches 186; Indels 122; Gaps 15;

```

QY 171 GCHDSSESSVVCVCPAGFTGSRCEHSOALHCHPEACPDATCVNRPDGRGYTCRCHLGR 230
Db 1 GTCTAIKADSVICLCLPLGF-----19
QY 231 SGLRCEGVTVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS---283
Db 20 KGRHCEDAFTLTIQFRESLSRYAATPWLEPQHYLSFMFEIIFRPSDGGVLLYSYDT 79
QY 284 GKSGPVEDFVSLAMVGGHLEFRYELGSLAVLSRAEPLALGRWHRVSAERLNKDGSLRV 343
Db 80 GSK-----DFLSINLAGGHVEFRFDGSGTGVLRSEDPDLTGNWHELVRSTAKNGILQV 134
QY 344 NGRPVLRSPPKSGQLNHLTLVLLGGVEPSVPLSPATNMSAHRGCVGEVSNKRLDL 403
Db 135 DKQKIVEGMAEGGFTQIKNTDIFIGGVNDDVKNKSGVLKPFSGSIQKIILNDRTHV 194
QY 404 TYSFLSGOIGQCYSDFPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCHEENPCOL 463
Db 195 KHDF--TSGVNVENAAPHCVRAPCAHGGSCRPEK-GYDCDCPLGFE-----239
QY 464 REPCLHGTCQGT--RCLCLPFGSPRCQSGHGIAESDWHLESGSGNDAPGQYGFYH 521
Db 240 ----LH---CQKAIIEAIEIPIQFIG-----RSYLTYN 265
QY 522 DDGFLAFPHGFVSRSLPEVPEITIEVTRTASGLLLWQGVVEGAGQKDFISLGLQDG 581
Db 266 PDILKRVSG---SRS-----NVFMRFTTAKDGLLWRG--DSPMRPNSDFISLGLRDG 314
QY 582 HLVPFYQLGSGEARLVSEDPINDGEWHRVTLALREGRRGSIQVDGEEILVSGRSPGNVAVN 641
Db 315 ALVFSYNLGGVASIMVNGSFNDGRWHRVKAVRDQSGKITVDDYGARTGKSPGMWRQLN 374
QY 642 AKGSVYIGGAPDVATLTGGRSSGITCVKKNLVLHSARPGAPPQPPDLQHRQAAGANTR 701
Db 375 INGALYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DYHISLVEDAVDGKVIN 427
QY 702 PC 703
Db 428 TC 429
  
```

RESULT 10
 US-10-073-865-78
 ; Sequence 78, Application US/10073865
 ; Publication No. US20030044904A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; CURRENT APPLICATION NUMBER: US/10/073,865
 ; CURRENT FILING DATE: 2002-02-14
 ; Prior Application removed - See file wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 154
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 78
 ; LENGTH: 432

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-073-865-78

Query Match 15.8%; Score 605; DB 14; Length 432;
Best Local Similarity 29.0%; Pred. No. 4.5e-41;
Matches 157; Conservative 75; Mismatches 188; Indels 122; Gaps 15;

QY 171 GCHDSSESSVVCVCPAGFTGSRCEHSOALHCHPEACPDATCVNRPDGRGYTCRCHLGR 230
Db 1 GTCTAIKADSVICLCLPLGF-----19
QY 231 SGLRCEGVTVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS---283
Db 20 KGRHCEDAFTLTIQFRESLSRYAATPWLEPQHYLSFMFEIIFRPSDGGVLLYSYDT 79
QY 284 GKSGPVEDFVSLAMVGGHLEFRYELGSLAVLSRAEPLALGRWHRVSAERLNKDGSLRV 343
Db 80 GSK-----DFLSINLAGGHVEFRFDGSGTGVLRSEDPDLTGNWHELVRSTAKNGILQV 134
QY 344 NGRPVLRSPPKSGQLNHLTLVLLGGVEPSVPLSPATNMSAHRGCVGEVSNKRLDL 403
Db 135 DKQKIVEGMAEGGFTQIKNTDIFIGGVNDDVKNKSGVLKPFSGSIQKIILNDRTHV 194
QY 404 TYSFLSGOIGQCYSDFPCERQPCQHGATCMPAGEYEFQCLCRDGFKGDLCHEENPCOL 463
Db 195 KHDF--TSGVNVENAAPHCVRAPCAHGGSCRPEK-GYDCDCPLGFE-----239
QY 464 REPCLHGTCQGT--RCLCLPFGSPRCQSGHGIAESDWHLESGSGNDAPGQYGFYH 521
Db 240 ----LH---CQKAIIEAIEIPIQFIG-----RSYLTYN 265
QY 522 DDGFLAFPHGFVSRSLPEVPEITIEVTRTASGLLLWQGVVEGAGQKDFISLGLQDG 581
Db 266 PDILKRVSG---SRS-----NVFMRFTTAKDGLLWRG--DSPMRPNSDFISLGLRDG 314
QY 582 HLVPFYQLGSGEARLVSEDPINDGEWHRVTLALREGRRGSIQVDGEEILVSGRSPGNVAVN 641
Db 315 ALVFSYNLGGVASIMVNGSFNDGRWHRVKAVRDQSGKITVDDYGARTGKSPGMWRQLN 374
QY 642 AKGSVYIGGAPDVATLTGGRSSGITCVKKNLVLHSARPGAPPQPPDLQHRQAAGANTR 701
Db 375 INGALYVGMKEIALHTNRQYMRGLVGCISHFTLST-----DYHISLVEDAVDGKVIN 427
QY 702 PC 703
Db 428 TC 429
  
```

RESULT 11
 US-10-242-747-102
 ; Sequence 102, Application US/10242747
 ; Publication No. US20040005577A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PT207C1
 ; CURRENT APPLICATION NUMBER: US/10/242,747
 ; CURRENT FILING DATE: 2002-09-13
 ; PRIOR APPLICATION NUMBER: 09/764,881
 ; PRIOR FILING DATE: 2001-01-17
 ; PRIOR APPLICATION NUMBER: 60/179,065
 ; PRIOR FILING DATE: 2000-01-31
 ; PRIOR APPLICATION NUMBER: 60/180,628
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: 60/214,886
 ; PRIOR FILING DATE: 2000-06-28
 ; PRIOR APPLICATION NUMBER: 60/217,487
 ; PRIOR FILING DATE: 2000-07-11
 ; PRIOR APPLICATION NUMBER: 60/225,758
 ; PRIOR FILING DATE: 2000-08-14
 ; PRIOR APPLICATION NUMBER: 60/220,963
 ; PRIOR FILING DATE: 2000-07-26

; PRIOR APPLICATION NUMBER: 60/217,496
 ; PRIOR FILING DATE: 2000-07-11
 ; PRIOR APPLICATION NUMBER: 60/225,447
 ; PRIOR FILING DATE: 2000-08-14
 ; PRIOR APPLICATION NUMBER: 60/218,290
 ; PRIOR FILING DATE: 2000-07-14
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 192
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 102
 ; LENGTH: 432
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-242-747-102

Query Match 15.8%; Score 605; DB 15; Length 432;
 Best Local Similarity 29.0%; Pred. No. 4.5e-41;
 Matches 157; Conservative 75; Mismatches 188; Indels 122; Gaps 15;

QY 171 GOCHDESSYVCVCPAGFTGSRCEHSQALHCHPEACGPDATCVARNPDGRGYTCRCHLGR 230
 Db 1 GTCTAIKADSYICLCPGIF-----
 QY 231 SGLRCEGVTVTPSL-SGAGSYLALPALNTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
 Db 20 KGRHCEDAFTLITPQFRESLRSYAATFWPELPQHYLSFMEFEITFRPDSGDGVLVLYDYT 79
 QY 284 GKSGPVEDVSLAMVGGHLEFRYELGSLAVLSAEPALGRWHRVSAERLNKDGSLRV 343
 Db 80 GSK-----DPLSNLAGHVEFRPDCGSGTVLRSBEPDLTGNWHELRSVRAKGILOY 134
 QY 344 NGRPVLRSPPGKSGQLNHTLLYLGGVFPVPLSPATNMSAHRFCVGEVSWGKRLDL 403
 Db 135 DKQKIVEGMAEGGFTQIKCNTDIFIGGVNPDVVKKNSGLVKLPFGSGSIKIIINDRTIHW 194
 QY 404 TYSFLSQGIGQCYDSSPCERQCGATCMPAGEYEFQCLCEGDKGKDLCHHEENPQOL 463
 Db 195 KHF--TSGVNVENAAHPCVRAEPCAHGGSCRPEK-GYDCDCPLGPEG----- 239
 QY 464 REPCRHGTCQGT--RCLCLPFGSPRCQGGSGHGAESDWHLEGGNDAPGQYCAVYFH 521
 Db 240 ---LH---CQKAIIBAIPIQIG-----RSYLYTDN 265
 QY 522 DDGFLAFPGHVFRSRLPEVPEETIELEVRTSTASGILLWQVGEAGCGQKDFISLQDQG 581
 Db 266 PDILKRVSG---SRS-----NVFMRFKTKADKGLLWRG--DSPMRNSDFISLGRDQ 314
 QY 582 HLVFRYQLGSGEARLYSEDPINDGEMHVRTALREGRGSIQVDGEBELVSGRSPGNVAVN 641
 Db 315 ALVFSYNLGSGVAVSINVNGSFNDGRWHRVKAVDGQSGKITYDDYCARTKGKSPGMMRQLN 374
 QY 642 AKGSVYIGGAPDVATLTGGRFSSGTTGCVKNNLVLSHARPGAPPQPLDLOHRAQAQANTR 701
 Db 375 INGALYVGGMKEIATLNRQYMRGLVGCISHFTLST-----DVHISLIVEDAVDGGKIN 427
 QY 702 PC 703
 Db 428 TC 429

RESULT 12
 US-10-016-283-25
 ; Sequence 25, Application US/10016283
 ; Publication No. US20020164702A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela et al., David M.
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
 ; FILE REFERENCE: REG195-B-PCT-US
 ; CURRENT APPLICATION NUMBER: US/10/016,283
 ; CURRENT FILING DATE: 2001-11-30
 ; PRIOR APPLICATION NUMBER: US/09/077,955A
 ; PRIOR FILING DATE: 1998-09-10
 ; PRIOR APPLICATION NUMBER: PCT/US96/20696
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 26
 ; LENGTH: 440
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-016-283-26

; PRIOR FILING DATE: 1996-12-13
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 25
 ; LENGTH: 456
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-016-283-25

Query Match 15.2%; Score 579.5; DB 13; Length 456;
 Best Local Similarity 32.7%; Pred. No. 6e-39;
 Matches 168; Conservative 57; Mismatches 156; Indels 133; Gaps 16;

QY 242 TTPSLSGAGSYLALPALNTNTHHEL-----LDVEFKPLAPDGVLLFSGSKS 287
 Db 25 TASQDQSGPFLA-DFNFGSHLELRGLHTFARDIGERKVALELVFVFLARGPSGLLYNGOKT 83
 QY 288 GPVDFVSLAMVGGHLEFRYELGSLAVLSAEPALGRWHRVSAERLNKDGSLRVNNGR 347
 Db 84 DGKDFVSLALRRDRLEFRYDLGKGAAVIRSREPVTLGAWTRVSLERNRKRKALRVGDGP 143
 QY 348 PVLRSPPGKSGQ-----LNHTLLYLGGVFPVPLSPATNMSAHRFCVGEVSWGKRL- 401
 Db 144 RVLGESP-KSRKVPHITVNLKPELYVGGAPDFSKLARAAAAVSSGFDGAIQLVSLGSRQLL 202
 QY 402 -----DLTYSFLSQGIGQCYDSSPCER---QPCQHGATCMPAGEYEFQCLCRDGF 449
 Db 203 TPEHVLRAQVDT-SFAG-----HPCTRASGHEFLNGASCVP----- 237
 QY 450 KGDLCHEHENPCQUREPCLHGTCQGTFRCLCPFGSPRCQGGSGHGAESDWHLEGGSG 509
 Db 238 -----REAA-----VCLCPGSGFSGHCEKGL-----VEKSAG 265
 QY 510 NDAPGQYGVFHDGFLAFPHVFRSRLPEVPE-----IELEVRTSTA 553
 Db 266 -----DVTLLAFDGRFVFLVAVTSELANEIPEVKALQSNFELSRLTEAT 313
 QY 554 SGLLWQVGEVGEAGQKDFISLQDGHVFRYQLGSGEARLYSEDPINDGEMHVRTAL 613
 Db 314 QGLVMS---GKATERADYVALAIVDGHLQLSINLGSQPWVLRSTVAVNTNRWLRVVAH 369
 QY 614 REGRGSIQVDGEBELVSGRSPGNVAVNAGSYVIGGAPD--VATLTGGRFSSGTTGCVK 671
 Db 370 REQREGSLQVNEAPVTVGSSPLGATQDITDGLMWLGLPELVPGLPKAYGTGTVGCLR 429
 QY 672 NLVLSHARPGAPPQPLDLOHRAQAQANTRPCPS 705
 Db 430 DVVVGR-----HPLHLEDAVTKPELRPCPT 455

RESULT 13
 US-10-016-283-26
 ; Sequence 26, Application US/10016283
 ; Publication No. US20020164702A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela et al., David M.
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
 ; FILE REFERENCE: REG195-B-PCT-US
 ; CURRENT APPLICATION NUMBER: US/10/016,283
 ; CURRENT FILING DATE: 2001-11-30
 ; PRIOR APPLICATION NUMBER: US/09/077,955A
 ; PRIOR FILING DATE: 1998-09-10
 ; PRIOR APPLICATION NUMBER: PCT/US96/20696
 ; PRIOR FILING DATE: 1996-12-13
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 26
 ; LENGTH: 440
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-016-283-26

Query Match 15.1%; Score 576; DB 13; Length 440;

```

Best Local Similarity 32.9%; Pred. No. 1.1e-38;
Matches 163; Conservative 55; Mismatches 145; Indels 132; Gaps 15;
QY 261 THLELR-----LDVEFKLAPDGVLLFSGKSGVEDFVSLAMVGGHLEFR 306
Db 27 SHLELRGLHTFARDLGEKMALEVFLARGSPGLLYNGQTDGKDFVSLALDRRLEFR 86
QY 307 YELGSGLAARSAEPALGRWHRYSARLNKDSLRVNGRPVLRSPKSG-----LN 361
Db 87 YDLGKGAIVKRSPEVTLGATWRVSLRNGRKGALRVGDGPRVLGESP-KSRKVPHTVLN 145
QY 362 LHLLYLGGVEPSPVLPATNMSAHFRGCVGEVSVNGKRL-----DLTYSFLGSO 411
Db 146 LKEFLYVGGAPDFSKLARAAAASSGFDAQLVSLGGRQLLTPHVLRLQVDVT-SFAG-- 202
QY 412 GIGOCYSSPCR---QPCQHGATCMPAGEYEFQCLCRDGFKGDLCHEHENPCQLRPCL 468
Db 203 -----HPCTRASGHPCLNGASCVP-----REAA 226
QY 469 HGGTCQGTCLCLPFGSPGRCQOQSGHGIAESDWHLEGGNDAPQGYGAYFDDGFLAF 528
Db 227 -----VCLCPGFGSPHCKEL-----VKSAG-- 256
QY 529 PGHFRSLSEVPET-----IELEVRTSTASGLLWQGVGEGAGQKD 572
Db 257 DGRTFEYLNAVTESELANEIPVEKALQSNHFELSRLTEATQGLVLS-----GKATERAD 312
QY 573 FISLQDGHVFPYQLGSGEARLVSDPINDGWHRTALREGRSIQVDGEEELVSGR 632
Db 313 YVALAVDGHLSYNGSPVLRSTVNTWRLRVVAHRQGRSGLVGNEAVTGS 372
QY 633 SPGNVAVNAKGSVYIGAPD--VATLTGGRFSSGITCVKKNLVLSHARSFAPPPQLDL 690
Db 373 SPLGATQDIDGALMLGGLPELPGPALPKAYGTGFVGLRDLVVVGR-----HPLHL 424
QY 691 QHRAQAGANTRPCS 705
Db 425 LEDAVTPELRPCPT 439

RESULT 14
US-09-978-249-10
; Sequence 10, Application US/09978249
; Patent No. US20020106780A1
; GENERAL INFORMATION:
; APPLICANT: Fiscella, et al.
; TITLE OF INVENTION: Extracellular Matrix Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT054P1
; CURRENT APPLICATION NUMBER: US/09/978,249
; PRIOR FILING DATE: 2001-10-17
; PRIORITY FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/198,123
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10
; LENGTH: 375
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-978-249-10

Query Match 14.2%; Score 543.5; DB 9; Length 375;
Best Local Similarity 30.6%; Pred. No. 4.1e-36;
Matches 136; Conservative 70; Mismatches 161; Indels 77; Gaps 12;
QY 265 LRLDVEFKLAPDGVLLFS---GKSGVEDFVSLAMVGGHLEFRYELGSGLAARSAEP 321
Db 1 MEFEITRPDSGDVLLYSYDTSK-----DFLSINLAGHVFREFDCSGTGLRSED 55
QY 322 LALGRWHRYSARLNKDSLRVNGRPVLRSSPKKQGLNLHLLYLGGVEPSPVLRPAT 381
Db 56 LTJGNWHELVRSTAKNGIQLQVDKQKIVGEMAGSFTQIKCNVDIFIGVFNVDVKNK 115
    
```

```

QY 382 NMSAHFRGCVGEVSVNGKRLDLTVSFLGSGIGOCYDSSPCRQPCQGATCMPAGEYEF 441
Db 116 GYLKFEFSGSIQKILINDRTIHKHDF--TSGVNVENAAHPVCRAPCARHGGSCRRPKB-GY 172
QY 442 OCLCRDRGFKGDLCEHEENPCQLRPCLHGGTCQGT--RCLCLPGFSGPRCOQSGHGIAE 499
Db 173 DDCPLGEG-----LH---CQKAITEAIEIPIQFIG----- 200
QY 500 SDWHLEGGSGNDAPQGYGAYFDDGFLAFPGHVFVSRSLPEVETIELEVRTSTASGLLW 559
Db 201 -----RSYLTYDNPDIKRVSG---SR-----NVFMRFKITAKDGLLW 237
QY 560 QGVVEAGQKDTISIGLQDGHVFPYQLGSGEARLVSDPINDGWHRTALREGSRG 619
Db 238 RG--DSPMRPNSDFTISLGRDGLVFSYNLGGVASIMVNGSFNDGRWHRVKAIVDQSG 295
QY 620 SIQVDGEBELVSGRSPGNVAVNAKGSVYIGAPDVAITLTGGRFSSGITCVKKNLVLSH 679
Db 296 KITVDDYGARTGKSPGMWRQLNINGALYVGMKEIALHTNRQYMRGLVGCISHFTLSI-- 353
QY 680 PEAAPPQDLQHRAQAGANTRPC 703
Db 354 -----DYHISLVEDAVDGNKINTC 372

RESULT 15
US-10-016-283-27
; Sequence 27, Application US/10016283
; Publication No. US20020164702A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/10/016,283
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US/09/077,955A
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: PCT/US96/20696
; PRIOR FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 27
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-283-27

Query Match 13.8%; Score 528; DB 13; Length 390;
Best Local Similarity 33.0%; Pred. No. 8.1e-35;
Matches 149; Conservative 50; Mismatches 134; Indels 118; Gaps 14;
QY 291 EDFVSLAMVGGHLEFRYELGSGLAARSAEPALGRWHRYSARLNKDSLRVNGRPVL 350
Db 21 DDKYSALALRRRLRLEFRYDLGKGAIVRSREPVTLGATWRVSLRNGRKGALRVGDGPRVL 80
QY 351 RSPKSGQ-----LNLHLLYLGGVEPSPVLPATNMSAHFRGCVGEVSVNGKRL--- 401
Db 81 GESP-KSRKVPHTVNLKPELPLVGGAPDFSKLARAAAASSGFDAQLVSLGGRQLTPE 139
QY 402 -----DLTYSFLGSGIGOCYDSSPCR---QPCQHGATCMPAGEYEFQCLCRDRGFKGD 452
Db 140 HVLRQVDVT-SFAG-----HPCTRASGHPCLNGASCVP----- 171
QY 453 LCEHEENPCQLRECLHGGTCQGTFCCLCPGFSRPRCOQSGHGIAESDWHLEGGGND 512
Db 172 -----REAA-----VCLCPGFGSPHCKEL-----VKSAG-- 199
QY 513 PQGYGAYFDDGFLAFPGHVFVSRSLPEVET-----IELEVRTSTASGL 556
Db 200 -----DVIDTLAFDGRTFEYLNAVTESELANEIPVEKALQSNHFELSRLTEATQGL 250
QY 557 LLWQGVGEGAGQKDFISLQDGHVFPYQLGSGEARLVSDPINDGWHRTALREG 616
    
```

```

  251 VLMS---GKATERADYVALAIVDGHLLQLSINLGSQFVLRSTVFVNTNRWLRVAHREQ 306
  617 RRSIQVDGEBELVSRSPGNVAVNAKGSVYICGAPD--VATLTGGRFSSGITGCVKNLV 674
  307 REGSLOVGNBAPVTGSSPLGATQLDITDGLWLGELPELVGYPALPKAYGTGFVGCCLRDVY 366
  675 LHSARPGAPPPQPLDQHRACAGANTPCPS 705
  367 VGR-----HPLLHLEDAVTKPELRPCPT 389
  
```

Search completed: March 9, 2004, 17:25:14
 Job time : 54.7431 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:16:23 ; Search time 27.6355 Seconds
(without alignments)
1317.011 Million cell updates/sec

Title: US-10-006-011A-3
Perfect score: 3825
Sequence: 1 EIKITFRPSADGMLLYNGQ.....QPLDLQHRQAQANTRPCPS 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	862.5	22.5	1940	2	US-08-644-271-30
2	862.5	22.5	1940	4	US-09-077-955-34
3	640.5	16.7	492	2	US-08-644-271-32
4	640.5	16.7	492	4	US-09-077-955-36
5	579.5	15.2	456	4	US-09-077-955-25
6	576	15.1	440	4	US-09-077-955-26
7	528	13.8	390	4	US-09-077-955-27
8	397.5	10.4	771	3	US-09-188-930-183
9	397.5	10.4	1525	3	US-09-191-647-2
10	397.5	10.4	1525	3	US-09-540-245A-2
11	397.5	10.4	1525	3	US-08-340-153-2
12	394.5	10.3	1529	4	US-09-312-283C-396
13	394	10.3	716	4	US-09-312-283C-183
14	391.5	10.2	1523	4	US-09-182-024A-2
15	382	10.0	338	4	US-09-077-955-28
16	361	9.4	3075	2	US-08-460-309-5
17	361	9.4	3075	2	US-08-125-077-5
18	359	9.4	1150	6	5444158-2
19	357	9.3	1130	2	US-08-460-309-2
20	357	9.3	1130	2	US-08-125-077-2
21	357	9.3	3088	4	US-09-562-702A-8
22	357	9.3	3089	4	US-09-562-702A-4
23	357	9.3	3110	4	US-09-562-702A-2
24	357	9.3	3110	4	US-08-562-702A-7
25	357	9.3	3110	4	US-09-561-709B-6
26	357	9.3	3111	2	US-08-460-309-4
27	357	9.3	3111	2	US-08-125-077-4

28	349.5	9.1	735	3	US-09-191-647-9
29	349.5	9.1	735	3	US-09-540-245A-9
30	349.5	9.1	735	3	US-09-540-153-9
31	347	9.1	3084	4	US-09-562-702A-12
32	347	9.1	3106	4	US-09-562-702A-10
33	340	8.9	294	4	US-09-077-955-29
34	333	8.7	1480	3	US-09-191-647-7
35	333	8.7	1480	3	US-09-540-245A-7
36	333	8.7	1480	3	US-09-540-153-7
37	333	8.7	1480	5	PCT-US91-09055-2
38	328.5	8.6	2471	1	US-08-185-432-16
39	328.5	8.6	2471	1	US-08-083-590A-19
40	328.5	8.6	2471	3	US-08-532-384-19
41	328.5	8.6	2471	4	US-08-899-322-1
42	328	8.6	1480	4	US-08-182-024A-5
43	326	8.5	256	4	US-09-077-955-30
44	322.5	8.4	3571	4	US-09-911-842A-2
45	316	8.3	2321	4	US-09-230-652-2

ALIGNMENTS

RESULT 1
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/644,271
FILING DATE: 10-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 60/008,657
FILING DATE: 15-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 195A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721
TELEX:
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 1940 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Rat Agrin
LOCATION: 1..1940
OTHER INFORMATION:
US-08-644-271-30

Query Match 22.5% ; Score 862.5 ; DB 2 ; Length 1940;

Best Local Similarity 32.0%; Pred. No. 7.4e-64; Indels 135; Gaps 22; Matches 239; Conservative 92; Mismatches 280;

2 IKITFRPDSADGMLLYNQKRVGSPPTLANRQPDFISFLVGGRRPFRFDAGSGMATIR 61
1287 LALFRALTEGSLLYNGNA-----RGKDFLALALLDGRVQFRFDTSGGPAVIT 1335

62 HPTPLALGHFHTVILLRSLTQGLSIVGDLAPVNGTSQKFGQLDNEELYLGGVPDYG-- 119
1336 SLVPVPPGRWHRLELSRHWROGTLSDGETPVVGGSPSGTDGLNLDNLNYVGGIPEEQVA 1395

120 -AIPKAGLSSGFIGVRELRIOGEEIVFHDNLNTA---HGISHC--PTCRDRPCQNGQC 173
1396 MVDRTSVGVLKGCIRMLDINNQQLELSDWQRAAVQSSVGGECGDHPCLNPNCHGGALC 1455

174 HDSSESSVYVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGL 233
1456 QALEAGMFLCQCPFRFGPTCADEKS-PCQNPCHGAAPCRVLSGG-GAKCECPLGRSGT 1513

234 RCEEGVTTTPSLSGAG-----SYLALPALTNWHEHL---RLDVEKFLPADG 278
1514 FCQ---TV----LETAGSRPFLADFNFGFSYLELKGHTFERDLGKMALEWVFLARQPSG 1566

279 VLLFSGGKGFVDFVSLAMVGGHLEFRYELGSLGLAVLSAEPALALGRWHRVSAERLNKD 338
1567 LLLYNGQKTDGKDFVSLALNHRHLEFCYDLGKGAAVIRSKKEPIALGTWVRFLENRGRK 1626

339 GSLRVNGRPPVLRSSPGKSQG----LNHLTLLYLGVEPSPVPLSATNWSAHPGCVGE 393
1627 GALQVGDGPRVIGESP-KSRKVPHTMLNKEFLYIGGAPDFSKLARGAAVSSGFSGVIOQL 1685

394 VSVNGKRDLTAVSFLGSGIGOCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGD 453
1686 VSLRGHQL-----LTQEHVLRADVSPFADHPCTQA----- 1716

454 CEHENPCLREPCHLHGTC---QGT-RCLCLPGRSPRCCQSGHGIAESDWHLEGGSG 509
1717 -----LGNPCLNGGSCVPREATYECCLPCGFSGLHCEKGL-----VEKSVG 1757

510 NDAPGOYGAFHDDGFLAPPGHVRSRSLPEVPET-----IELEVRTSTASGLLLWQG 561
1758 -----DLETFLAPDGRTYEYLNVAIESEKALQSNHFELSLRTEATQGLVLM-- 1803

562 VEVGAGGKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEHWRTALREGRRSI 621
1804 --IGKAAERADYMALAIVDGHLQLSVDLGSQPVVLRSTVKVNTNRMLRIRAHREHREGSL 1861

622 QVDGEBELYSRSPGNVAVNAKGSVYIGGAP--DVAVITLGGFRSSGITGCKVNLVHLSAR 679
1862 QVGNRAPVTVGSSFLGATQDIDGALWGLGLOKLPVGOALPKAYGTGFCCLRDVVVGH-- 1919

680 PGAPPPQPLDLOHRAQAGANTRPCPS 705
1920 -----RQLHLEDVAVTKPELRPCTP 1939

RESULT 2
US-09-077-955-34
Sequence 34, Application US/09077955A
Patent No. 6413740

GENERAL INFORMATION:
APPLICANT: Valenzuela et al., David M.
TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS

FILE REFERENCE: REG195-B-PCT-US
CURRENT APPLICATION NUMBER: US/09/077,955A
CURRENT FILING DATE: 1998-09-10

EARLIER APPLICATION NUMBER: PCT/US96/20696
EARLIER FILING DATE: 1996-12-13
EARLIER APPLICATION NUMBER: 08/644,271

EARLIER FILING DATE: 1996-05-10
EARLIER APPLICATION NUMBER: 60/008,657
EARLIER FILING DATE: 1995-12-15
NUMBER OF SEQ ID NOS: 36

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 34
LENGTH: 1940
TYPE: PRT
ORGANISM: Rattus sp.
US-09-077-955-34

Query Match 22.5%; Score 862.5; DB 4; Length 1940;
Best Local Similarity 32.0%; Pred. No. 7.4e-64;
Matches 239; Conservative 92; Mismatches 280; Indels 135; Gaps 22;

2 IKITFRPDSADGMLLYNQKRVGSPPTLANRQPDFISFLVGGRRPFRFDAGSGMATIR 61
1287 LALFRALTEGSLLYNGNA-----RGKDFLALALLDGRVQFRFDTSGGPAVIT 1335

62 HPTPLALGHFHTVILLRSLTQGLSIVGDLAPVNGTSQKFGQLDNEELYLGGVPDYG-- 119
1336 SLVPVPPGRWHRLELSRHWROGTLSDGETPVVGGSPSGTDGLNLDNLNYVGGIPEEQVA 1395

120 -AIPKAGLSSGFIGVRELRIOGEEIVFHDNLNTA---HGISHC--PTCRDRPCQNGQC 173
1396 MVDRTSVGVLKGCIRMLDINNQQLELSDWQRAAVQSSVGGECGDHPCLNPNCHGGALC 1455

174 HDSSESSVYVCPAGFTGSRCEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGRSGL 233
1456 QALEAGMFLCQCPFRFGPTCADEKS-PCQNPCHGAAPCRVLSGG-GAKCECPLGRSGT 1513

234 RCEEGVTTTPSLSGAG-----SYLALPALTNWHEHL---RLDVEKFLPADG 278
1514 FCQ---TV----LETAGSRPFLADFNFGFSYLELKGHTFERDLGKMALEWVFLARQPSG 1566

279 VLLFSGGKGFVDFVSLAMVGGHLEFRYELGSLGLAVLSAEPALALGRWHRVSAERLNKD 338
1567 LLLYNGQKTDGKDFVSLALNHRHLEFCYDLGKGAAVIRSKKEPIALGTWVRFLENRGRK 1626

339 GSLRVNGRPPVLRSSPGKSQG----LNHLTLLYLGVEPSPVPLSATNWSAHPGCVGE 393
1627 GALQVGDGPRVIGESP-KSRKVPHTMLNKEFLYIGGAPDFSKLARGAAVSSGFSGVIOQL 1685

394 VSVNGKRDLTAVSFLGSGIGOCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKGD 453
1686 VSLRGHQL-----LTQEHVLRADVSPFADHPCTQA----- 1716

454 CEHENPCLREPCHLHGTC---QGT-RCLCLPGRSPRCCQSGHGIAESDWHLEGGSG 509
1717 -----LGNPCLNGGSCVPREATYECCLPCGFSGLHCEKGL-----VEKSVG 1757

510 NDAPGOYGAFHDDGFLAPPGHVRSRSLPEVPET-----IELEVRTSTASGLLLWQG 561
1758 -----DLETFLAPDGRTYEYLNVAIESEKALQSNHFELSLRTEATQGLVLM-- 1803

562 VEVGAGGKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEHWRTALREGRRSI 621
1804 --IGKAAERADYMALAIVDGHLQLSVDLGSQPVVLRSTVKVNTNRMLRIRAHREHREGSL 1861

622 QVDGEBELYSRSPGNVAVNAKGSVYIGGAP--DVAVITLGGFRSSGITGCKVNLVHLSAR 679
1862 QVGNRAPVTVGSSFLGATQDIDGALWGLGLOKLPVGOALPKAYGTGFCCLRDVVVGH-- 1919

680 PGAPPPQPLDLOHRAQAGANTRPCPS 705
1920 -----RQLHLEDVAVTKPELRPCTP 1939

RESULT 3
US-08-644-271-32
Sequence 32, Application US/08644271
Patent No. 5814478

GENERAL INFORMATION:
APPLICANT: Valenzuela, et al.
TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS

TITLE OF INVENTION: AND LIGANDS
NUMBER OF SEQUENCES: 32

```

CORRESPONDENCE ADDRESS:
ADDRESS: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/644,271
FILING DATE: 10-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 60/008,657
FILING DATE: 15-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 195A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721

```

```

INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
FEATURES:
NAME/KEY: Human Agrin
LOCATION: 1..492
OTHER INFORMATION:
US-08-644-271-32

```

```

Query Match      16.7%; Score 640.5; DB 2; Length 492;
Best Local Similarity 33.0%; Pred. No. 6.6e-46;
Matches 183; Conservative 59; Mismatches 173; Indels 139; Gaps 18;

QY 202 CHPACGPDATCVNRPDGRGYTCRCHLGRSGLRCEBGTVTTFPSLSGAGSYLALPALANT 261
Db 27 CQPNPCHGAAPCRVLPFG-GAQCECPLEGRTFCQ-----TAGQDGGSPFLA-DFNGFS 79

QY 262 HHELR-----LDVEFKPLADPQVLLFFSGKSGPVEDFVSLAMVGGHLEFRY 307
Db 80 HLELRGLHTFARDLGEKMALEVVLARGPSGILLYNGQKTGKGFVSLALDRDRELEFRY 139

QY 308 ELGSLAVLRSABPLALGRWHRVSAERLNKDGSLRVNNGRPNLRSSPKSKSQG-----LNL 362
Db 140 DLGKGAAVRSREPVTLGAWTRVSLERNGKALRVGDGPRVLTGESP-KSRKVPHTVLNL 198

QY 363 HTLLYLGGVEPSPVLSPATNMSAHFRGCVEVSNQKRL-----DLTYSFLLSGSQG 412
Db 199 KEPLYVGGAPDFSKLARAASVSSGFDGAIQLVSLGGRLQTPHVLRODVT-SPAG- 254

QY 413 IGQCYDSSPCER--OPCOHGATCMPAGEYEFQCLCRDGFKDCHEHENPCQLREPCILH 469
Db 255 -----HPCTRASGHFCLNAGSCVP-----REAY- 278

QY 470 GGTCQGRTRCLLPGFSGRCCQSGHGIAESDWHLEGGNGNDAPQGYGFHDDGFLAPF 529
Db 279 -----VCLCPGFGSPHCCKGL-----VEKSAG-----DVDTLAFD 309

QY 530 GHVFSRSLPEVPT-----IELEVRTSTASGLLLWQGVGEAGQGGKDF 573
Db 310 GRTEVEYLNNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLMW-----GKATERADY 365

QY 574 ISLGLDQHLVFRYQVLSGEGEARLVSEDPINDGEWHVRTALREGRGSIQVDGELVSGRS 633
Db 366 VALAIVDGHLOQLSYNLGSGQPVLVRSTVFNTRNLRVVAHREGRGSLQVGNAPVITGSS 425

```

```

QY 574 ISLGLDQHLVFRYQVLSGEGEARLVSEDPINDGEWHVRTALREGRGSIQVDGELVSGRS 633
Db 366 VALAIVDGHLOQLSYNLGSGQPVLVRSTVFNTRNLRVVAHREGRGSLQVGNAPVITGSS 425

QY 634 PGNVAVNAKSVYIGGAPD--VATLTGGRFSSGITCVKNIHLHSARPGAPPQPDLDLQ 691
Db 426 PLGATQDLDTGALWIGGPELPFVGFALPKAYGTGFCVGLKRDVVVGR-----HPLHL 477

QY 692 HRAQAGANTRPCPS 705
Db 478 EDVATKPELRCPPT 491

RESULT 4
US-09-077-955-36
; Sequence 36, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 36
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-077-955-36

```

```

Query Match      16.7%; Score 640.5; DB 4; Length 492;
Best Local Similarity 33.0%; Pred. No. 6.6e-46;
Matches 183; Conservative 59; Mismatches 173; Indels 139; Gaps 18;

QY 202 CHPACGPDATCVNRPDGRGYTCRCHLGRSGLRCEBGTVTTFPSLSGAGSYLALPALANT 261
Db 27 CQPNPCHGAAPCRVLPFG-GAQCECPLEGRTFCQ-----TAGQDGGSPFLA-DFNGFS 79

QY 262 HHELR-----LDVEFKPLADPQVLLFFSGKSGPVEDFVSLAMVGGHLEFRY 307
Db 80 HLELRGLHTFARDLGEKMALEVVLARGPSGILLYNGQKTGKGFVSLALDRDRELEFRY 139

QY 308 ELGSLAVLRSABPLALGRWHRVSAERLNKDGSLRVNNGRPNLRSSPKSKSQG-----LNL 362
Db 140 DLGKGAAVRSREPVTLGAWTRVSLERNGKALRVGDGPRVLTGESP-KSRKVPHTVLNL 198

QY 363 HTLLYLGGVEPSPVLSPATNMSAHFRGCVEVSNQKRL-----DLTYSFLLSGSQG 412
Db 199 KEPLYVGGAPDFSKLARAASVSSGFDGAIQLVSLGGRLQTPHVLRODVT-SPAG- 254

QY 413 IGQCYDSSPCER--OPCOHGATCMPAGEYEFQCLCRDGFKDCHEHENPCQLREPCILH 469
Db 255 -----HPCTRASGHFCLNAGSCVP-----REAY- 278

QY 470 GGTCQGRTRCLLPGFSGRCCQSGHGIAESDWHLEGGNGNDAPQGYGFHDDGFLAPF 529
Db 279 -----VCLCPGFGSPHCCKGL-----VEKSAG-----DVDTLAFD 309

QY 530 GHVFSRSLPEVPT-----IELEVRTSTASGLLLWQGVGEAGQGGKDF 573
Db 310 GRTEVEYLNNAVTESELANEIPVEKALQSNHFELSLRTEATQGLVLMW-----GKATERADY 365

QY 574 ISLGLDQHLVFRYQVLSGEGEARLVSEDPINDGEWHVRTALREGRGSIQVDGELVSGRS 633
Db 366 VALAIVDGHLOQLSYNLGSGQPVLVRSTVFNTRNLRVVAHREGRGSLQVGNAPVITGSS 425

```

```

QY 634 PGNVAVNAKGSVYIGGAPD--VATLTGGRFFSSITGCVKXNVLHSARPAPPDLDLQ 691
Db 426 PLGATOLDTDGALWGLGPELVPFALPKAYGTGFVGLRDVVVGR-----HPLHL 477
QY 692 HRAQAGANTRCPCS 705
Db 478 EDVATKPELRCPCT 491

```

```

RESULT 5
US-09-077-955-25
; sequence 25, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; EARLIER FILING DATE: 1998-09-10
; EARLIER FILING DATE: 1996-12-13
; EARLIER FILING DATE: 1996-05-10
; EARLIER FILING DATE: 1996-05-10
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-25

```

```

Query Match 15.2%; Score 579.5; DB 4; Length 456;
Best Local Similarity 32.7%; Pred. No. 8.8e-41;
Matches 168; Conservative 57; Mismatches 156; Indels 133; Gaps 16;

```

```

QY 242 TTPSLGAGSYALPALPNTTHELR-----LDVEPKPLAPDGVLLFSGCKS 287
Db 25 TASGQDGGPFLA-DFNFSHLELRGLHTFARDLGEKMALEVFLARGPSGLLYNQKT 83
QY 288 GPVEDFVSLAMVGGHLEFRYELSGSLAVLRSAPLALGRVHRSABERLNKDGSLRVNGGR 347
Db 84 DKGDFVSLARDRLRLEFRYDYGKAAVIRSEPVTLGAWTRVSLERNGRKGALRVDGDP 143
QY 348 PVLRSPPKSKQG-----LNLHTLVLGGEVPSPLSPATNMSAHFRGCVGEVSNGRKL- 401
Db 144 RVLGESP-KSRKVPHTVNLKREPLYVGGAPPDFSLARAAA VSSGFDGAIQLVSLGGRQLL 202
QY 402 -----DLTYSFLGSGGIGCCYDSSPCER---QPCHGATCMPAGEYEFQCLCRDGF 449
Db 203 TPEHVLRQVDVT-SFAG-----HPCTRASGHPCLINGASCVP----- 237
QY 450 KGDLCHEENPCQLREPLHGGTCCGTRCLCLPGFSGPRCOOGSGHGIABSDMHLGSGG 509
Db 238 -----REAY-----VCLCPGGFSGPHCEKGL-----VEKSAG 265
QY 510 NDPAGQYGFHDDGFLAPPGHVFRSPLPEVPEP-----IELEVRTSTASGLLWQGV 553
Db 266 -----DVTLDLAFDGRTFVEYLNVAVTESELANEIVPEKALQSNHPELSLRTEAT 313
QY 554 SGLLWQGVVEGAGQKDFISLGLQDGHVFRYQLGSGEARLYSEDPINDGEVHRYTAL 613
Db 314 QGLVLWMS-----GKATERYADYVALAVDGHQLSILNLSQPVVLRSTVTVNTRVLRVVAH 369
QY 614 REGRRGSIQVDGEEELVSGRSPGNVAVNAKGSVIYIGGAPD--VATLTGGRFFSSITGCVK 671
Db 370 REOREGSLQVNEAPVTGSSPLGATOLDTDGALWGLGPELVPFALPKAYGTGFVGLR 429
QY 672 NLVLSHARPCAPPDLDLQHRAQAGANTRCPCS 705
Db 430 DVVWGR-----HPLHLLLEDVATKPELRCPCT 455

```

```

RESULT 6
US-09-077-955-26
; Sequence 26, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; EARLIER FILING DATE: 1998-09-10
; EARLIER FILING DATE: 1996-12-13
; EARLIER FILING DATE: 1996-05-10
; EARLIER FILING DATE: 1996-05-10
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-26

```

```

Query Match 15.1%; Score 576; DB 4; Length 440;
Best Local Similarity 32.9%; Pred. No. 1.6e-40;
Matches 163; Conservative 55; Mismatches 145; Indels 132; Gaps 15;
QY 261 THELR-----LDVEPKPLAPDGVLLFSGCKSPVEDFVSLAMVGGHLEFR 306
Db 27 SHLELRGLHTFARDLGEKMALEVFLARGPSGLLYNQKTDKGDFVSLARDRLRLEFR 86
QY 307 YELSGSLAVLRSAPLALGRVHRSABERLNKDGSLRVNGGRPVLRSSPCKSQG-----LN 361
Db 87 YDLGKAAVIRSEPVTLGAWTRVSLERNGRKGALRVDGPRVLSGESP-KSRKVPHTVNL 145
QY 362 LHTLVLGGEVPSPLSPATNMSAHFRGCVGEVSNGRKL-----DLTYSFLGSG 411
Db 146 LKREPLYVGGAPPDFSLARAAA VSSGFDGAIQLVSLGGRQLLTPEHVLRQVDVT-SFAG-- 202
QY 412 GIGCCYDSSPCER---QPCHGATCMPAGEYEFQCLCRDGFKGDLCHEENPCQLREPL 468
Db 203 -----HPCTRASGHPCLINGASCVP-----REAY 226
QY 469 HGGTCCGTRCLCLPGFSGPRCOOGSGHGIABSDMHLGSGGNADPQYCAVPHDDGFLAF 528
Db 227 -----VCLCPGGFSGPHCEKGL-----VEKSAG-----DVTDLAF 256
QY 529 PGHVFRSPLPEVPEP-----IELEVRTSTASGLLWQGVVEGAGQGD 572
Db 257 DGRTFVEYLNVAVTESELANEIVPEKALQSNHPELSLRTEATQGLVWMS---GKATERAD 312
QY 573 FISLGLQDGHVFRYQLGSGEARLYSEDPINDGEVHRYTALREGRRGSIQVDGEEELVSGR 632
Db 313 YVALAIVDGHQLSILNLSQPVVLRSTVTVNTRVLRVVAHREOREGSLQVNEAPVTGS 372
QY 633 SPGNVAVNAKGSVIYIGGAPD--VATLTGGRFFSSITGCVKXNVLHSARPAPPDLDL 690
Db 373 SPLGATOLDTDGALWGLGPELVPFALPKAYGTGFVGLRDVVVGR-----HPLHL 424
QY 691 QHRAQAGANTRCPCS 705
Db 425 LEDAVTKPELRCPCT 439

```

```

RESULT 7
US-09-077-955-27
; Sequence 27, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS

```

```

; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; EARLIER FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 27
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-27

```

```

Query Match 13.8%; Score 528; DB 4; Length 390;
Best Local Similarity 33.0%; Pred. No. 1.6e-36;
Matches 149; Conservative 50; Mismatches 134; Indels 118; Gaps 14;

```

```

QY 291 EDFVSLAMVGGHLEFEYELGSLAVLRSAPLALGEWHRVSAERLNKDGSLRVNGRPVL 350
Db 21 DKVSLALDRRLLEFRYDLGKGAAVIRSRPEVTLGAWTRVSLERNRKGALRVGDGPRVL 80
QY 351 RSPGKSGQ-----LMLHTLLYLGGVPEPSVPLSPATNMSAHFRGCVGVSVNGKEL----- 401
Db 81 GESP-KSRKVPHTVNLKPELVGGAPDPSKLARAAVSSGFDGAIQLVSLGGRLLTPE 139
QY 402 -----DLTYSLGSGOICQYDSSPCR---QPCHGATCMPAGEYEFQCLCRDGFKGD 452
Db 140 HVLQVDVT-SFAG-----HPCTRASGHPCLNGASCVP----- 171
QY 453 LCEHEENPCQLREPCPLHGTCQCTRCCLLPFGSPRCOOGSGHGAESDWHLEGGGND 512
Db 172 -----REAY-----VLCFPGSGFHCXKGL-----VERKAG--- 199
QY 513 PQGYAYFDDGFLAPPHVFRSRLPEVPET-----LELVRTSTASGL 556
Db 200 -----DVTFLAFDGRTFEYLNVAVTESELANEIVFKALQSNHPELSLRTEATQGL 250
QY 557 LLWQGVGEAGQKDFSLGQDGHVFRYQLGSGEARLVSEDPINDGWHRYVTLREG 616
Db 251 VLWS---GKATERADYVALAVDGHQLSLYMLGSPVVLASVTVVNTWRWLRVVAHQ 306
QY 617 RGSIQVDGEEELVSRGSPFNVA NAKGSVYTGAPD--VATLTGCRFSSGTCVKNLV 674
Db 307 REGSLQVNEAPVTGSSPLGATQDLDGALWLGGLPELVPVGPALPKAYGTGTVGCLRDVV 366
QY 675 LHSARPGPPPOPLDQHRQAQAGNTRCPS 705
Db 367 VGR-----HPLHLLDEAVTKPELRPCPT 389

```

```

RESULT 8
US-09-188-930-183
; Sequence 183, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 183
; LENGTH: 771

```

```

; TYPE: PRT
; ORGANISM: Rat
; FEATURES:
; NAME/KEY: UNSURE
; LOCATION: (717)....(717)
US-09-188-930-183

```

```

Query Match 10.4%; Score 397.5; DB 3; Length 771;
Best Local Similarity 27.6%; Pred. No. 5e-25;
Matches 145; Conservative 64; Mismatches 185; Indels 131; Gaps 28;

```

```

QY 158 HC-----PTCDRPCQNGGQCHDSESSYVCVAGFTGSRCEHSQAL-----HCRPEAC 207
Db 315 HCDIDFDCCQNKCKGAHCTDA-VNGYTCVCEGYSGLFCEFSPPMVFURTPCQNFDC 373
QY 208 GPDATC---VNRPDGRGYTCRHLGRSGLRCBEGVITVTPSLSGAGSYLALPAL-----T 259
Db 374 QNGAQCIIRVNEP-----ICQCLPGYLGEKELVSVSI-LVNKESYLQIPSAKVRPQT 426
QY 260 NTHHELRLDVEFKPLADPDGVLFLSFGKSGVPEDFVSLAMVGGHLEFRYELGSLA-VLRS 318
Db 427 NITLQIATDED-----SGILLYKDK-----DHIAVESIEG-IRASYDTGSHPASAIYS 474
QY 319 AEPLALGRHRSVAERLNKDGSLRVNGRPEVLRSSPKSQCLNHLHTLLYLGGVE-----P 373
Db 475 VETINDGNFHI VELLTLDSSLSLSVDGSGPKIITNLSKQSTLNFDSPLLYVGGMPGKNWA 534
QY 374 SVPLSPATNMSAHFRGCVGVSVNGKRLDLYTSLFSGOIGQCYDSSPCRQPCQHGATC 433
Db 535 SLRQAPQGN-GTSFHCIRNLXINSBLQDFRKYPMQTGILPGC---EPCHKKYCAHG-TC 589
QY 434 MPAGEYEFQCLCRDGFKGDICEHEEN-PCQLREPCPLHGTCQ-----GTRCCLCPGFSGP 487
Db 590 QPSSQSGFTCEBEGWMLCQDRTDPC-LGNKCVH-GTCLPINAIFYSCCKLEHGCV 647
QY 488 RCQGGSHGIAESDWHLEGGGNDAPQGYGAYPHDDGFLAPPG-HVFSRSLVPEPETIEL 546
Db 648 LCDE-----EEDLFPNLPQDQVQAR----- 667
QY 547 EVRTSTASGLLW--OGVEVGEAGQKDFSLGQDGHVFRYQLGSGEARLVSEDPIND 604
Db 668 EVQALWARAALLWQWVHRGQLTQ-----RISCRGERIDYQ--SSVRCLL-----ND 716
QY 605 GEHWRVTA-----LREGR-----GSIQVDGEEELV 629
Db 717 XEVSRLCEKRCGACGCGCGLRASKRKYSECTDSSSFVDEVEKV 761

```

```

RESULT 9
US-09-191-647-2
; Sequence 2, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: 598-031-3
; CURRENT APPLICATION NUMBER: US/09/191,647
; CURRENT FILING DATE: 1998-11-13
; EARLIER APPLICATION NUMBER: 60/065,544
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: 60/081,057
; EARLIER FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-191-647-2

```

Query Match 10.4%; Score 397.5; DB 3; Length 1525;
 Best Local Similarity 29.7%; Pred. No. 1.3e-24;
 Matches 108; Conservative 58; Mismatches 141; Indels 57; Gaps 18;

Qy 158 HC-----PTCRDRPCQNGQCHDSESSSYVCVCPAGFTGSRCEHSQAL-----HCHPEAC 207
 Db 1068 HCDIDFDDCCDNKCKNGAHCTDA-VNGYTCICPEGYSGLFCFSPMWLPRTPSPCDNFD 1126
 Qy 208 GPDATC---VNRPDGRGYTCRHLGRSGLRCEBEGVTTTFLSGAGSYLALPAL-----T 259
 Db 1127 QNGAQCIVRINEP-----ICQCLPGYQGEKELVSV---NFINKESYLQIPSAKVRPQT 1178
 Qy 260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGEVEFVSLAMVGGHLEFRYELGSLA-VLRS 318
 Db 1179 NITLIQIATDED-----SGILLYKGDK-----DHIAVELYRGRVRSYDTGSHPASAIYS 1227

Qy 319 AEPLALGRWHRVSRABRLNKGDSLRVNGRVPVLRSSPKSQGLNHLHTLLYLGVE-----P 373
 Db 1228 VETINDGNFHIVELLALDQSLSLSDVGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVA 1287

Qy 374 SVPLSPATNMSAHRFGCVGVSVNKKRLDLYSFLSGOGIGQCVDSPCEPQCHGATC 433
 Db 1288 SLRQAPQGN-GTSPHGICIRNLYINSELQDFQKVPMTGILPGC---EPCHKVCAHG-TC 1342

Qy 434 MPAGEYFQCLCRDGFKGDICEHEEN-PCQLRBPCLHGHTCQ-----GTRCLCLPFGSPG 487
 Db 1343 QPSSQAGFTCECQEGWMPGLCDORTNDPC-LGNKCVH-GTCLPINAFAFSYCKLEHGCV 1400

Qy 488 RCOQ 491
 Db 1401 LCDE 1404

RESULT 10
 US-09-540-245A-2
 ; Sequence 2, Application US/09540245A
 ; Patent No. 6270984
 ; GENERAL INFORMATION:
 ; APPLICANT: Goodman, Corey
 ; APPLICANT: Kid, Thomas
 ; APPLICANT: Brose, Katja
 ; APPLICANT: Tessier-Lavigne, Marc
 ; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
 ; FILE REFERENCE: B98-031-3
 ; CURRENT APPLICATION NUMBER: US/09/540,245A
 ; CURRENT FILING DATE: 2000-03-31
 ; PRIOR APPLICATION NUMBER: 60/065,544
 ; PRIOR FILING DATE: 1997-11-14
 ; PRIOR APPLICATION NUMBER: 60/081,057
 ; PRIOR FILING DATE: 1998-04-07
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 1525
 ; TYPE: PRT
 ; ORGANISM: human
 US-09-540-245A-2

Query Match 10.4%; Score 397.5; DB 3; Length 1525;
 Best Local Similarity 29.7%; Pred. No. 1.3e-24;
 Matches 108; Conservative 58; Mismatches 141; Indels 57; Gaps 18;

Qy 158 HC-----PTCRDRPCQNGQCHDSESSSYVCVCPAGFTGSRCEHSQAL-----HCHPEAC 207
 Db 1068 HCDIDFDDCCDNKCKNGAHCTDA-VNGYTCICPEGYSGLFCFSPMWLPRTPSPCDNFD 1126
 Qy 208 GPDATC---VNRPDGRGYTCRHLGRSGLRCEBEGVTTTFLSGAGSYLALPAL-----T 259
 Db 1127 QNGAQCIVRINEP-----ICQCLPGYQGEKELVSV---NFINKESYLQIPSAKVRPQT 1178
 Qy 260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGEVEFVSLAMVGGHLEFRYELGSLA-VLRS 318
 Db 1179 NITLIQIATDED-----SGILLYKGDK-----DHIAVELYRGRVRSYDTGSHPASAIYS 1227

Qy 319 AEPLALGRWHRVSRABRLNKGDSLRVNGRVPVLRSSPKSQGLNHLHTLLYLGVE-----P 373
 Db 1228 VETINDGNFHIVELLALDQSLSLSDVGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVA 1287

Qy 374 SVPLSPATNMSAHRFGCVGVSVNKKRLDLYSFLSGOGIGQCVDSPCEPQCHGATC 433
 Db 1288 SLRQAPQGN-GTSPHGICIRNLYINSELQDFQKVPMTGILPGC---EPCHKVCAHG-TC 1342

Qy 434 MPAGEYFQCLCRDGFKGDICEHEEN-PCQLRBPCLHGHTCQ-----GTRCLCLPFGSPG 487
 Db 1343 QPSSQAGFTCECQEGWMPGLCDORTNDPC-LGNKCVH-GTCLPINAFAFSYCKLEHGCV 1400

Qy 488 RCOQ 491
 Db 1401 LCDE 1404

RESULT 11
 US-09-540-153-2
 ; Sequence 2, Application US/09540153
 ; Patent No. 6270995
 ; GENERAL INFORMATION:
 ; APPLICANT: Goodman, Corey
 ; APPLICANT: Kid, Thomas
 ; APPLICANT: Brose, Katja
 ; APPLICANT: Tessier-Lavigne, Marc
 ; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
 ; FILE REFERENCE: B98-031-3
 ; CURRENT APPLICATION NUMBER: US/09/540,153
 ; CURRENT FILING DATE: 2000-03-31
 ; PRIOR APPLICATION NUMBER: 09/191,647
 ; PRIOR FILING DATE: 1998-11-13
 ; PRIOR APPLICATION NUMBER: 60/081,057
 ; PRIOR FILING DATE: 1998-04-07
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 1525
 ; TYPE: PRT
 ; ORGANISM: human
 US-09-540-153-2

Query Match 10.4%; Score 397.5; DB 3; Length 1525;
 Best Local Similarity 29.7%; Pred. No. 1.3e-24;
 Matches 108; Conservative 58; Mismatches 141; Indels 57; Gaps 18;

Qy 158 HC-----PTCRDRPCQNGQCHDSESSSYVCVCPAGFTGSRCEHSQAL-----HCHPEAC 207
 Db 1068 HCDIDFDDCCDNKCKNGAHCTDA-VNGYTCICPEGYSGLFCFSPMWLPRTPSPCDNFD 1126
 Qy 208 GPDATC---VNRPDGRGYTCRHLGRSGLRCEBEGVTTTFLSGAGSYLALPAL-----T 259
 Db 1127 QNGAQCIVRINEP-----ICQCLPGYQGEKELVSV---NFINKESYLQIPSAKVRPQT 1178
 Qy 260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGEVEFVSLAMVGGHLEFRYELGSLA-VLRS 318
 Db 1179 NITLIQIATDED-----SGILLYKGDK-----DHIAVELYRGRVRSYDTGSHPASAIYS 1227

Qy 319 AEPLALGRWHRVSRABRLNKGDSLRVNGRVPVLRSSPKSQGLNHLHTLLYLGVE-----P 373
 Db 1228 VETINDGNFHIVELLALDQSLSLSDVGGNPKIITNLSKQSTLNFDSPLYVGGMPGKSNVA 1287

Qy 374 SVPLSPATNMSAHRFGCVGVSVNKKRLDLYSFLSGOGIGQCVDSPCEPQCHGATC 433
 Db 1288 SLRQAPQGN-GTSPHGICIRNLYINSELQDFQKVPMTGILPGC---EPCHKVCAHG-TC 1342

Qy 434 MPAGEYFQCLCRDGFKGDICEHEEN-PCQLRBPCLHGHTCQ-----GTRCLCLPFGSPG 487
 Db 1343 QPSSQAGFTCECQEGWMPGLCDORTNDPC-LGNKCVH-GTCLPINAFAFSYCKLEHGCV 1400

Qy 488 RCOQ 491
 Db 1401 LCDE 1404

```

Db      1401 LCDE 1404
RESULT 12
US-09-312-283C-396
; Sequence 396, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated from Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 396
; LENGTH: 1529
; TYPE: PRT
; ORGANISM: Mouse
US-09-312-283C-396

Query Match      10.3%; Score 394.5; DB 4; Length 1529;
Best Local Similarity 30.2%; Pred. No. 2,4e-24;
Matches 110; Conservative 54; Mismatches 143; Indels 57; Gaps 18;

QY 158 HC-----PTCRDRPCONGQCHDSSESSYVCVCPAGFTGSRCEHSQAL-----HCHPEAC 207
Db 1072 HCDIDFDQDNKCKNGAHCCTDA-VNGYTCVCEGYSGLFCFSPMVLRTSPCDNFDC 1130
QY 208 GPDATC---VNRDGRGYTCRCHLGRSGLRCEGVTVTTPSLGAGSYLALPAL-----T 259
Db 1131 QNGAQCIIRVNEP-----ICQCLPGLYLGKCKELVSVI---NFVKESYLQIPSAKVRPQT 1182
QY 260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSLA-VLRS 318
Db 1183 NITLQIATDED-----SGILLYKGDK-----DHIAVESIEG-IRASYDTGSHPASAIYS 1231
QY 319 APELALGRHRYSAERLNKDGSLRVNGRPLVRSRSPKSGQLNLHTLLYLGVE-----P 373
Db 1232 VETINDGNFHVHLLTDLSSLSVDGSGPKITNLKQSTLNFDFSPLYVGGMPGKNNVA 1291
QY 374 SVPLSPATNMSAHRFCVCGEVSVNGKRLDLYSFLGSGQIGQCYDSSPCERQPCQHGATC 433
Db 1292 SLRQAPGQN-GTSFHGCIIRNLINSELDQFRKVPMTGILPGC---EPCHKVCAHG-TC 1346
QY 434 MPAGEYFQCLCRDGFKGDLCHEHEN-PCQLREPCVLHGTCQ-----GTRCLCLPGFSGP 487
Db 1347 QPSSQSGFTCEBEGMGLCPDQRTNDPC-LGNKCVH-GTCLPINAFAFSYCKLEGRHGGV 1404
QY 488 RCOQ 491
Db 1405 LCDE 1408

RESULT 13
US-09-312-283C-183
; Sequence 183, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated from Skin Cells
; TITLE OF INVENTION: and Methods for Their Use

```

```

; FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 183
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Mouse
US-09-312-283C-183

Query Match      10.3%; Score 394; DB 4; Length 716;
Best Local Similarity 28.3%; Pred. No. 8,8e-25;
Matches 126; Conservative 59; Mismatches 160; Indels 100; Gaps 23;

QY 158 HC-----PTCRDRPCONGQCHDSSESSYVCVCPAGFTGSRCEHSQAL-----HCHPEAC 207
Db 315 HCDIDFDQDNKCKNGAHCCTDA-VNGYTCVCEGYSGLFCFSPMVLRTSPCDNFDC 373
QY 208 GPDATC---VNRDGRGYTCRCHLGRSGLRCEGVTVTTPSLGAGSYLALPAL-----T 259
Db 374 QNGAQCIIRVNEP-----ICQCLPGLYLGKCKELVSVI---LVNKSYLEIIPSAKVRPQT 426
QY 260 NTHHELRLDVEFKPLAPDGVLLFSGGKSGPVEDFVSLAMVGGHLEFRYELGSLA-VLRS 318
Db 427 NITLQIATDED-----SGILLYKGDK-----DHIAVESIEG-IRASYDTGSHPASAIYS 474
QY 319 APELALGRHRYSAERLNKDGSLRVNGRPLVRSRSPKSGQLNLHTLLYLGVE-----P 373
Db 475 VETINDGNFHVHLLTDLSSLSVDGSGPKITNLKQSTLNFDFSPLYVGGMPGKNNVA 534
QY 374 SVPLSPATNMSAHRFCVCGEVSVNGKRLDLYSFLGSGQIGQCYDSSPCERQPCQHGATC 433
Db 535 SLRQAPGQN-GTSFHGCIIRNLINSELDQFRKVPMTGILPGC---EPCHKVCAHG-TC 589
QY 434 MPAGEYFQCLCRDGFKGDLCHEHEN-PCQLREPCVLHGTCQ-----GTRCLCLPGFSGP 487
Db 590 QPSSQSGFTCEBEGMGLCPDQRTNDPC-LGNKCVH-GTCLPINAFAFSYCKLEGRHGGV 647
QY 488 RCOQSGHGAESDWHLEGGSGNDAPQYGAFFDDGFLAFPG-HVFSRSLRPEVPTIEL 546
Db 648 LCDE-----EEDLNFPLPGDQVQAR----- 667
QY 547 EVRTSTASGILLW--QGEVEVGEAQ 569
Db 668 EVOALWARAALLWMOQWIHRGOLTO 692

RESULT 14
US-09-182-024A-2
; Sequence 2, Application US/09182024A
; Patent No. 6342370
; GENERAL INFORMATION:
; APPLICANT: Connolly, Timothy
; APPLICANT: Rajput, Bhanu
; TITLE OF INVENTION: Human Slit Polypeptide and Polynucleotides Encoding
; TITLE OF INVENTION: Same
; FILE REFERENCE: 640100-271
; CURRENT APPLICATION NUMBER: US/09/182,024A
; CURRENT FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 60/063,946
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/096,420
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1523
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-182-024A-2

Query Match      10.2%; Score 391.5; DB 4; Length 1523;

```

Best Local Similarity 29.0%; Pred. No. 4.3e-24;
 Matches 129; Conservative 53; Mismatches 158; Indels 105; Gaps 24;

QY 133 CVRELRIQGE---EIVFH---DLNLTAR-----GIS-----HCPY-----CR 163
 DB 1020 CFCPPNTYGLCEVIDHCVPKLNQHEAKCIPDKGFCCECVGSGKLCETDNDCCV 1079
 QY 164 DRPCONGSOCHDSSSSVVCVCPAGFTGSRCEHS-----QALHCHPEACGPPDATCV--- 214
 DB 1080 AHKCRHGAQCVDIT-INGYTCPCPGFSGFPCHEHPPMVLQTSFCDQYECQNGAQCIVVQ 1138
 QY 215 NRDPGRGYTRCHLGRGLRCEBEGVTTPFSLGAGSYLALPALINTHRLRDLDFEKL 274
 DB 1139 QEP-----TCRCPFGAPRCEKLTIV---NFVKDSYVEL--ASAKVRQANISLQVATD 1189
 QY 275 APDGVILFSGKSGSPVEDFVSLAMVGHLEFRYE-LSGLAVLRSAPLALGRHWRVSAE 333
 DB 1190 KDNIGILLYKGD-----NDPLALELYQGHVRLVYDVSNSPPTTVYSVETVNDGGQFHSVEVV 1244
 QY 334 RLKNDGSLRVNGRPFVLRSSPKSQGLNHLTLVGLGVSPSPVLSPATNMS-----AHFRG 389
 DB 1245 TLNQTLNLVWDKGTPKSLGRFKQPAVGINSPYLGIPITSTGLSALRQGTDRPLGGFHG 1304
 QY 390 CVGEVSNKRLD---LTYFSLG-SQIGQC-----YDSSPCERQP----- 426
 DB 1305 CIHEVRINNELQDFKALFPOSLSVSPGCKSTVCKHGLCRSVEKDSVVCRECFGTGPLC 1364
 QY 427 -----COHGATCMPAGEYEFQCLCRDGFKGDLCHEEE---NPCOLREPCLLHG 471
 DB 1365 DQEARDFCLHRCHHG-KCVATGT-SYMKCKAEGYGDLDCKNDSANACSAPK--CHHG 1420
 QY 472 TC-----QG-TRCLCLPGFSGPRCQ 491
 DB 1421 QCHISDQGEPCYCLCQPGFSGHCCQ 1445

RESULT 15
 US-09-077-955-28
 ; Sequence 28, Application US/09077955A
 ; Patent No. 6413740
 ; GENERAL INFORMATION:
 ; APPLICANT: Valenzuela et al., David M.
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
 ; FILE REFERENCE: REG195-B-PCT-US
 ; CURRENT APPLICATION NUMBER: US/09/077,955A
 ; CURRENT FILING DATE: 1998-09-10
 ; EARLIER APPLICATION NUMBER: PCT/US96/20696
 ; EARLIER FILING DATE: 1996-12-13
 ; EARLIER APPLICATION NUMBER: 08/644,271
 ; EARLIER FILING DATE: 1996-05-10
 ; EARLIER APPLICATION NUMBER: 60/008,657
 ; EARLIER FILING DATE: 1995-12-15
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 28
 ; LENGTH: 338
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-077-955-28

Query Match 10.0%; Score 382; DB 4; Length 338;
 Best Local Similarity 29.6%; Pred. No. 3.1e-24;
 Matches 118; Conservative 42; Mismatches 125; Indels 114; Gaps 13;

QY 338 DGSRLVNGRPFVLRSSPKSQGLNHLTLVGLGVSPSPVLSPATNMSAHFRGCVGEVSNV 397
 DB 22 DKGPRVILGESPKSRKVP--HTVNLNKEPLYVGGAPDFSKLARAAAASVSSGFDGAIQLVSLG 79
 QY 398 GKRL-----DLTYFSLGSGIGQCYDSSPCER---OPCOHGATCMPAGEYEFQCL 444
 DB 80 GRQLLTPEHVLRQVDVT-SFAG-----HFCTRASGHPCINGASCYP----- 119
 QY 445 CRDFGKGLDCEHEENPQLREPCLLHGTCQTRCLCLPGFSGPRCQGGSHGIAESDWHL 504

DB 120 -----REAAAY-----VCLCPGFGSPHCEKGL-----V 142
 QY 505 EGSNGNDAPGOYGAYFDDGFLAFFFVHVSRSLSPEVPEP-----IELEV 548
 DB 143 EKSAG-----DVTDLAFDGRFTVEYLNNAVTESELANEIPVEKALQSNHFELS 190
 QY 549 RTSTASGILLWQGVVEGAGQKDFISLGLDGHVFRYQLGSGEARLVSEDDPINDGBWH 608
 DB 191 RREATQGLVUMS-----GRATERADYVALAVDGHLLQLSYNLGSQFVVLURSTVFNTRWL 246
 QY 609 RVTALRREGRRGSIQVDGEBELVSGRSPGNVAVNAKGSVYIIGGAPD--VAITLGGFRFSSGI 666
 DB 247 RVVAHREQREGSLQVNEAPVTGSSPLGATQDIDGALWLGLELPELVGPPALPKAYGTGF 306
 QY 667 TCCYKXVILVHSARPCAPPPOPLDLOHRAQAQANTRPCPS 705
 DB 307 VGCLRDVVVGR-----HPLHLLLEDVATKPELRPCPT 337

Search completed: March 9, 2004, 17:23:32
 Job time: 29.6355 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:11:22 ; Search time 85.8465 Seconds
(without alignments)
2320.373 Million cell updates/sec

Title: US-10-006-011A-3
Perfect score: 3825
Sequence: 1 EIKITRPDSADQMLLYNQ.....QFLDLQHRAGANTRPCFS 705

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_29Jan04:**
- 1: Geneseqp1980s:**
 - 2: Geneseqp1990s:**
 - 3: Geneseqp2000s:**
 - 4: Geneseqp2001s:**
 - 5: Geneseqp2002s:**
 - 6: Geneseqp2003s:**
 - 7: Geneseqp2003Bs:**
 - 8: Geneseqp2004s:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3825	100.0	4391	6 AAE34390	Aae34390 Human per
2	3815	99.7	4392	4 AAB31889	Aab31889 Amino aci
3	3795	99.2	4436	4 ABG23265	Abg23265 Novel hum
4	1030	26.9	195	4 AAB31890	Aab31890 Amino aci
5	901.5	23.6	1931	6 ABUS2400	Abu52400 Human GPC
6	891.5	23.3	2053	7 ADC39154	Adc39154 Novel hum
7	849.5	22.2	2143	7 ADC39164	Adc39164 Novel hum
8	840	22.0	671	6 ABP58231	Abp58231 Human cel
9	840	22.0	1009	7 ADP28105	Adp28105 Human NTR
10	832	21.8	819	5 ADP72291	Adp72291 Rat prote
11	812.5	21.2	4072	4 ABB63614	Abb63614 Drosophi
12	664.5	17.4	1298	4 ABB61948	Abb61948 Drosophi
13	640.5	16.7	492	2 AAW26609	Aaw26609 Human agr
14	635.5	16.6	1566	7 ADC39166	Adc39166 Novel hum
15	628	16.4	463	7 ADP64904	Adp64904 Human pro
16	625.5	16.4	1565	7 ADC39156	Adc39156 Novel hum
17	605	15.8	432	4 AAU18102	Aau18102 Novel hum
18	605	15.8	432	4 AAU16938	Aau16938 Human nov
19	605	15.8	432	4 ABB10233	Abb10233 Human cDN
20	605	15.8	432	4 AAU19905	Aau19905 Novel hum
21	605	15.8	432	5 ABJ05729	Abj05729 Novel hum
22	605	15.8	432	5 ABP66820	Abp66820 Human pol
23	587.5	15.4	416	2 AAY73993	Aay73993 Human pro
24	576.5	15.1	1039	4 ABB63124	Abb63124 Drosophi
25	556	14.5	1741	5 ABP43859	Abp43859 Human mRN

26	543.5	14.2	375	4	AAU07421	Novel hum
27	441	11.5	3313	7	ADES4405	Rat Prote
28	441	11.5	3313	7	ADES4409	Rat Prote
29	420.5	11.0	1712	4	AAU28190	Novel hum
30	412.5	10.8	1327	4	AAW78871	Human pro
31	412	10.8	1522	4	AAW79855	Human pro
32	407	10.6	1508	2	AAZ27141	Human sli
33	407	10.6	1508	2	AAW96706	Protein s
34	407	10.6	1508	2	AAW96706	Protein s
35	407	10.6	1534	2	AAW4138	Human sli
36	407	10.6	1534	2	AAW46966	Amino aci
37	407	10.6	1534	2	AAZ27144	Human sli
38	407	10.6	1534	2	AAW96707	Protein s
39	407	10.6	1534	2	AAW96707	Protein s
40	401.5	10.5	1514	7	AAE54473	Human sli
41	398	10.4	1438	5	ADD49063	Human PRO
42	397.5	10.4	1503	2	AAE17600	Human ext
43	397.5	10.4	1503	2	AAZ27142	Human sli
44	397.5	10.4	1525	2	AAW96701	Sliit-like
45	397.5	10.4	1525	2	AAW96701	Sliit-like
46	397.5	10.4	1525	7	ADB85322	Human sli
47	397.5	10.4	1525	7	ADB85322	Human ner

ALIGNMENTS

RESULT 1
AAE34390
ID AAE34390 standard; protein; 4391 AA.
XX AC AAE34390;
XX DT 14-MAY-2003 (first entry)
DE Human perlecan protein.
KW Human; diagnosis; osteoarthritis; rheumatoid arthritis; perlecan.
XX OS Homo sapiens.
XX PN WO200295415-A2.
XX PD 28-NOV-2002.
XX PF 22-MAY-2002; 2002WO-EP005612.
XX PR 23-MAY-2001; 2001GB-00012626.
XX PA (OSTE-) OSTEOMETER BIO TECH AS.
XX PI Christgau S, Henriksen DB, Cloos PAC;
XX DR WPI; 2003-140389/13.
XX PT An assay for the diagnosis or assessment of the severity of
osteoarthritis or rheumatoid arthritis comprising detecting an isomerized
or optically inverted protein in a sample.
XX PS Disclosure; Page 46-67; 106pp; English.
XX CC The invention relates to an assay for the diagnosis or assessment of the
severity of osteoarthritis or rheumatoid arthritis. The assay involves
measuring (in a biological sample) the amount or presence of an
isomerized or optically inverted protein or one or more isomerized or
optically inverted fragments from proteins such as perlecan, biglycan,
decorin, fibrillin-1 or procollagen. The assay is useful for the
diagnosis or assessment of the severity of osteoarthritis or rheumatoid
arthritis. The present sequence is human perlecan protein

Query Match 100.0%; Score 3825; DB 6; Length 4391;
Best Local Similarity 100.0%; Pred. No. 2.6e-233;
Matches 705; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 EIKITFRPDSADGMLLYNGOKRVGSPNLANRQPDFISFGLVGRPEFRFDAGSGMATI 60
Db 3687 EIKITFRPDSADGMLLYNGOKRVGSPNLANRQPDFISFGLVGRPEFRFDAGSGMATI 3746
QY 61 RHPTLALGHFHTVLLRSLSLQGLSVGDLAPVNTGSKGFKGDLNBEELVLYGYPDYGA 120
Db 3747 RHPTLALGHFHTVLLRSLSLQGLSVGDLAPVNTGSKGFKGDLNBEELVLYGYPDYGA 3806
QY 121 IPKAGLSGFIGCVRELIQGEIVFVHDLNLTANGISHCPTCRDRPCQNGQCHDSSESS 180
Db 3807 IPKAGLSGFIGCVRELIQGEIVFVHDLNLTANGISHCPTCRDRPCQNGQCHDSSESS 3866
QY 181 YVCVCPAGFTGSRCHSHALHCHPEACGPDATCVNRPDGRGVTCTCHLGRSGLRCEBGT 240
Db 3867 YVCVCPAGFTGSRCHSHALHCHPEACGPDATCVNRPDGRGVTCTCHLGRSGLRCEBGT 3926
QY 241 VTPSLSGAGSYLALPALTNTTHELRDVEFKPLAPDGVLLFSGKSGPVEDFVSLAMVG 300
Db 3927 VTPSLSGAGSYLALPALTNTTHELRDVEFKPLAPDGVLLFSGKSGPVEDFVSLAMVG 3986
QY 301 GHLEFRYEIYELGSLAVLRSAPLGRWHRVAERLNKDGSLRVNNGRPRVLRSSPKSQGL 360
Db 3987 GHLEFRYEIYELGSLAVLRSAPLGRWHRVAERLNKDGSLRVNNGRPRVLRSSPKSQGL 4046
QY 361 NLHTLLYLGVEPPSPATNMSAHFRGCVGEVSNVNGKRLDLTYSEFLSGQIGOCYDSS 420
Db 4047 NLHTLLYLGVEPPSPATNMSAHFRGCVGEVSNVNGKRLDLTYSEFLSGQIGOCYDSS 4106
QY 421 PCERQPCOQGHGATCMPAGEYEFQCLCRDGFKDLCEHEENPCQLREPCLLHGTTCQGRCLC 480
Db 4107 PCERQPCOQGHGATCMPAGEYEFQCLCRDGFKDLCEHEENPCQLREPCLLHGTTCQGRCLC 4166
QY 481 LFGSGRCCQGGHGIAESDWHLEGGGNDAPQYGFYFDDDFLAFPPGHVFRSRLPEV 540
Db 4167 LFGSGRCCQGGHGIAESDWHLEGGGNDAPQYGFYFDDDFLAFPPGHVFRSRLPEV 4226
QY 541 PTELEVRTASGLLWQVGEAGQKDFISLQDGLHIVFYRLQSGEARLVSED 600
Db 4227 PTELEVRTASGLLWQVGEAGQKDFISLQDGLHIVFYRLQSGEARLVSED 4286
QY 601 PINDEWHRTALREGRRGSIQDGEELVSRGSPNPNVAVNAKGSVYIGAPDVATLTGG 660
Db 4287 PINDEWHRTALREGRRGSIQDGEELVSRGSPNPNVAVNAKGSVYIGAPDVATLTGG 4346
QY 661 RPSGITCVKLVHLSARPGAPPQLDLQHRQAQGANTRPCPS 705
Db 4347 RPSGITCVKLVHLSARPGAPPQLDLQHRQAQGANTRPCPS 4391

```

```

RESULT 2
AAB31889
ID AAB31889 standard; protein; 4393 AA.

```

```

XX
AC AAB31889;
DT 15-MAY-2001 (first entry)
DE Amino acid sequence of a human protein.
KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
OS Homo sapiens.
FN WO200105422-A2.
PD 25-JAN-2001.
PF 17-JUL-2000; 2000WO-FR002057.

```

```

XX
PA (INMR ) BIOMERIEUX STELHYS.
PI Roeklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
DR WPI; 2001-159475/16.
XX
PT Detecting, preventing and treating degenerative, neurological and
PT autoimmune diseases, particularly multiple sclerosis, using specified
PT polypeptides or related nucleic acid or ligand.
XX
PS Claim 1; Page 138-152; 209pp; French.
XX
CC The present sequence represents a human protein, which is used in the
CC method of the invention. The specification describes a method which uses
CC at least one polypeptide or polynucleotide sequence belonging to the
CC perlecan, precursor of the retinol-binding plasma protein, precursor of
CC the ganglioside GM2 activator, calgranulin B or saposin B protein
CC families. The method is used for detecting, preventing or treating a
CC degenerative, neurological and/or auto-immune disease. The
CC polynucleotides and polypeptides are used for diagnosis, prognosis,
CC prevention and treatment of multiple sclerosis (in its various forms and
CC phases). They may also be useful in cases of e.g. Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
CC polyarthritis and lupus erythematosus, including use as vaccines and in
CC gene therapy (expression of sense or antisense sequences). They can also
CC be used to assess efficacy of potential therapeutic agents, particularly
CC compounds that reduce or inhibit toxicity towards glial cells
XX
SQ Sequence 4393 AA;

```

```

Query Match 99.7%; Score 3815; DB 4; Length 4393;
Best Local Similarity 99.6%; Pred. No. 1.1e-232;
Matches 702; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 EIKITFRPDSADGMLLYNGOKRVGSPNLANRQPDFISFGLVGRPEFRFDAGSGMATI 60
Db 3689 EIKITFRPDSADGMLLYNGOKRVGSPNLANRQPDFISFGLVGRPEFRFDAGSGMATI 3748
QY 61 RHPTLALGHFHTVLLRSLSLQGLSVGDLAPVNTGSKGFKGDLNBEELVLYGYPDYGA 120
Db 3749 RHPTLALGHFHTVLLRSLSLQGLSVGDLAPVNTGSKGFKGDLNBEELVLYGYPDYGA 3808
QY 121 IPKAGLSGFIGCVRELIQGEIVFVHDLNLTANGISHCPTCRDRPCQNGQCHDSSESS 180
Db 3809 IPKAGLSGFIGCVRELIQGEIVFVHDLNLTANGISHCPTCRDRPCQNGQCHDSSESS 3868
QY 181 YVCVCPAGFTGSRCHSHALHCHPEACGPDATCVNRPDGRGVTCTCHLGRSGLRCEBGT 240
Db 3869 YVCVCPAGFTGSRCHSHALHCHPEACGPDATCVNRPDGRGVTCTCHLGRSGLRCEBGT 3928
QY 241 VTPSLSGAGSYLALPALTNTTHELRDVEFKPLAPDGVLLFSGKSGPVEDFVSLAMVG 300
Db 3929 VTPSLSGAGSYLALPALTNTTHELRDVEFKPLAPDGVLLFSGKSGPVEDFVSLAMVG 3988
QY 301 GHLEFRYEIYELGSLAVLRSAPLGRWHRVAERLNKDGSLRVNNGRPRVLRSSPKSQGL 360
Db 3989 GHLEFRYEIYELGSLAVLRSAPLGRWHRVAERLNKDGSLRVNNGRPRVLRSSPKSQGL 4048
QY 361 NLHTLLYLGVEPPSPATNMSAHFRGCVGEVSNVNGKRLDLTYSEFLSGQIGOCYDSS 420
Db 4049 NLHTLLYLGVEPPSPATNMSAHFRGCVGEVSNVNGKRLDLTYSEFLSGQIGOCYDSS 4108
QY 421 PCERQPCOQGHGATCMPAGEYEFQCLCRDGFKDLCEHEENPCQLREPCLLHGTTCQGRCLC 480
Db 4109 PCERQPCOQGHGATCMPAGEYEFQCLCRDGFKDLCEHEENPCQLREPCLLHGTTCQGRCLC 4168
QY 481 LFGSGRCCQGGHGIAESDWHLEGGGNDAPQYGFYFDDDFLAFPPGHVFRSRLPEV 540
Db 4169 LFGSGRCCQGGHGIAESDWHLEGGGNDAPQYGFYFDDDFLAFPPGHVFRSRLPEV 4228

```

QY 541 PETIELEVRTSTASGLLLWQVGEAGQCKDFISLGLQDHLVFRYQLGSGEARLVSBD 600
 |||||
 Db 4229 PETIELEVRTSTASGLLLWQVGEAGQCKDFISLGLQDHLVFRYQLGSGEARLVSBD 4288
 |||||
 QY 601 PINDGEWHRTALREGRGSIQVDGEBELVSGRPGPNVAVNAKGSYVIGGAPDVAITLGG 660
 |||||
 Db 4289 PINDGEWHRTALREGRGSIQVDGEBELVSGRPGPNVAVNAKGSYVIGGAPDVAITLGG 4348
 |||||
 QY 661 RFSSGITGCVKNIWLHSARFGAPPDLDLQHRAQAGANTRPCPS 705
 |||||
 Db 4349 RFSSGITGCVKNIWLHSARFGAPPDLDLQHRAQAGANTRPCPS 4393
 |||||

RESULT 3

ABG23265
 ID ABG23265 standard; protein; 4436 AA.
 XX AC
 XX ABG23265;

XX DT 18-FEB-2002 (first entry)
 XX DE

XX DE Novel human diagnostic protein #23256.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS
 XX XX Homo sapiens.
 XX XX W0200175067-A2.
 XX XX 11-OCT-2001.
 XX XX 30-MAR-2001; 2001WO-US008631.
 XX XX 31-MAR-2000; 2000US-00540217.
 XX XX 23-AUG-2000; 2000US-00649167.
 XX XX (HYSE-) HYSEQ INC.
 XX XX Drmanac RT, Liu C, Tang YT;
 XX XX WPI; 2001-639362/73.
 XX XX N-PSDB; AAS87452.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX PS
 XX PS Claim 20; SEQ ID NO 53624; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 4436 AA;
 Query Match 99.2%; Score 3795; DB 4; Length 4436;
 Best Local Similarity 98.2%; Pred. No. 2.1e-231;
 Matches 701; Conservative 1; Mismatches 2; Indels 10; Gaps 1;
 QY 1 EIKITFRPDSADGMLLVNQGKRVPGSPNLANRPDPFISGLVGRREFFRDAGSGMATI 60
 |||||
 Db 3722 EIKITFRPDSADGMLLVNQGKRVPGSPNLANRPDPFISGLVGRREFFRDAGSGMATI 3781
 |||||
 QY 61 RHPTPLALGHFHTVTLRLSLTQSSLIYVGDLPVNGTSGKFGQGLDNEELYLGGYPPYGA 120
 |||||
 Db 3782 RHPTPLALGHFHTVTLRLSLTQSSLIYVGDLPVNGTSGKFGQGLDNEELYLGGYPPYGA 3841
 |||||
 QY 121 IPKAGLSGGFVRELIQGEIIVFHDINLTAHGISHCPTCRDRPCQNGQCHDSESS 180
 |||||
 Db 3842 IPKAGLSGGFVRELIQGEIIVFHDINLTAHGISHCPTCRDRPCQNGQCHDSESS 3901
 |||||
 QY 191 YVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRDPDGRGYTCRHLRSGLRCEBGMT 240
 |||||
 Db 3902 YVCVPAGFTGSRCEHSQALHCHPEACGPDATCVNRDPDGRGYTCRHLRSGLRCEBGMT 3961
 |||||
 QY 241 VTTFSLGAGSYLALPALTWTHHELRLDVEFKLAPDGVLLFSGGKSGPVEDFVSLAMVG 300
 |||||
 Db 3962 VTTFSLGAGSYLALPALTWTHHELRLDVEFKLAPDGVLLFSGGKSGPVEDFVSLAMVG 4021
 |||||
 QY 301 GHLEFRYELGSLAVLRSABEPLALGRWHRVSAERLNKDGSLRVNNGRPPVLRSSPKSQGL 360
 |||||
 Db 4022 GHLEFRYELGSLAVLRSABEPLALGRWHRVSAERLNKDGSLRVNNGRPPVLRSSPKSQGL 4081
 |||||
 QY 361 NHTLLYLVGVEPSPVPLSPATNMSAHFRGCVGEVSNVNGKRLDLYSLGSGIGQCYDSS 420
 |||||
 Db 4082 NHTLLYLVGVEPSPVPLSPATNMSAHFRGCVGEVSNVNGKRLDLYSLGSGIGQCYDSS 4141
 |||||
 QY 421 PCERQPCOHGATCMPAGEYEFQCLCRDGFKGDICEHEENPCOLREPCCLHGGTCGTCLC 480
 |||||
 Db 4142 PCERQPCOHGATCMPAGEYEFQCLCRDGFKGDICEHEENPCOLREPCCLHGGTCGTCLC 4201
 |||||
 QY 481 LPFGSPRCQGGHGAESDWHLEGGNDAPQYGAYFHDGFLAFPGHVFSRSLPEV 540
 |||||
 Db 4202 LPFGSPRCQGGHGAESDWHLEGGNDAPQYGAYFHDGFLAFPGHVFSRSLPEV 4261
 |||||
 QY 541 PETIELEVRTSTASGLLLWQVGEAGQCKDFISLGLQDHLVFRYQLGSGEARLVSBD 590
 |||||
 Db 4262 PETIELEVRTSTASGLLLWQVGEAGQCKDFISLGLQDHLVFRYQLGSGEARLVSBD 4321
 |||||
 QY 591 SGEARLVSEDPINDGEWHRTALREGRGSIQVDGEBELVSGRPGPNVAVNAKGSYVIGG 650
 |||||
 Db 4322 SGEARLVSEDPINDGEWHRTALREGRGSIQVDGEBELVSGRPGPNVAVNAKGSYVIGG 4381
 |||||
 QY 651 APDVATLTGGRFSSGITGCVKNIWLHSARFGAPPDLDLQHRAQAGANTRPCP 704
 |||||
 Db 4382 APDVATLTGGRFSSGITGCVKNIWLHSARFGAPPDLDLQHRAQAGANTRPCP 4435
 |||||

RESULT 4

AAB31890

ID AAB31890 standard; protein; 195 AA.

XX AC AAB31890;

XX DT 15-MAY-2001 (first entry)

XX DE Amino acid sequence of the C-terminal of the human perlecan protein.

XX KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.

XX OS Homo sapiens.

XX XX

FN WO200105422-A2.
 XX 25-JAN-2001.
 XX 17-JUL-2000; 2000WO-FR002057.
 PF 15-JUL-1999; 99FR-00009372.
 PR (INNR) BIOMERIEUX STELHYS.
 PA Roecklin D, Kolibe H, Charles M, Malcus C, Santoro L, Perron H;
 PI WPI; 2001-159475/16.
 DR N-PSDB; AAF54728.
 XX
 XX Detecting, preventing and treating degenerative, neurological and
 PT autoimmune diseases, particularly multiple sclerosis, using specified
 PT polypeptides or related nucleic acid or ligand.
 XX
 XX Claim 1; Page 152-153; 209pp; French.
 XX
 CC The present sequence represents a human polypeptide, which is used in the
 CC method of the invention. The specification describes a method which uses
 CC at least one polypeptide or polynucleotide sequence belonging to the
 CC parlean, precursor of the retinol-binding plasma protein, precursor of
 CC the ganglioside GM2 activator, calgranulin B or saposin B protein
 CC families. The method is used for detecting, preventing or treating a
 CC degenerative, neurological and/or auto-immune disease. The
 CC polynucleotides and polypeptides are used for diagnosis, prognosis,
 CC prevention and treatment of multiple sclerosis (in its various forms and
 CC phases). They may also be useful in cases of e.g. Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and in
 CC gene therapy (expression of sense or antisense sequences). They can also
 CC be used to assess efficacy of potential therapeutic agents, particularly
 CC compounds that reduce or inhibit toxicity towards glial cells
 XX
 SQ Sequence 195 AA;
 Query Match 26.9%; Score 1030; DB 4; Length 195;
 Best Local Similarity 100.0%; Pred. No. 1.3e-57;
 Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 511 DAFQYGFHDDGFLAFGHVFSRSLPEVPEVTELEVTSTASGLLWQGVGEAGQG 570
 Db 1 DAFQYGFHDDGFLAFGHVFSRSLPEVPEVTELEVTSTASGLLWQGVGEAGQG 60
 QY 571 KDFISLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLRGRGSIQVDGEEELVS 630
 Db 61 KDFISLQDGHVFRYQLGSGEARLVSEDPINDGEWHRVTLRGRGSIQVDGEEELVS 120
 QY 631 GRSPGFNVAVNAGSVYIGAPDVALTGTGRFSSGITGVKNLVLHSRPPGAPPQPLDL 690
 Db 121 GRSPGFNVAVNAGSVYIGAPDVALTGTGRFSSGITGVKNLVLHSRPPGAPPQPLDL 180
 QY 691 QHRAQAGANTRPCPS 705
 Db 181 QHRAQAGANTRPCPS 195
 RESULT 5
 ABU52400
 ID ABU52400 standard; protein; 1931 AA.
 XX AC ABU52400;
 XX AC
 XX DT 03-MAR-2003 (first entry)
 XX DE Human GPCR related protein NOV40a.
 XX KW Human; NOVX; G-protein coupled receptor; GPCR; cancer; cyostatic.
 XX CS Homo sapiens.

XX WO200279398-A2.
 XX 10-OCT-2002.
 XX 08-MAR-2002; 2002WO-US007355.
 PR 08-MAR-2001; 2001US-0274194P.
 PR 08-MAR-2001; 2001US-0274281P.
 PR 09-MAR-2001; 2001US-0274322P.
 PR 13-MAR-2001; 2001US-0274849P.
 PR 13-MAR-2001; 2001US-0275578P.
 PR 13-MAR-2001; 2001US-0275579P.
 PR 14-MAR-2001; 2001US-0275601P.
 PR 16-MAR-2001; 2001US-0276000P.
 PR 19-MAR-2001; 2001US-0276776P.
 PR 20-MAR-2001; 2001US-0277239P.
 PR 20-MAR-2001; 2001US-0277327P.
 PR 21-MAR-2001; 2001US-0277388P.
 PR 21-MAR-2001; 2001US-0277791P.
 PR 22-MAR-2001; 2001US-0277833P.
 PR 23-MAR-2001; 2001US-0278152P.
 PR 26-MAR-2001; 2001US-0278894P.
 PR 27-MAR-2001; 2001US-0278999P.
 PR 27-MAR-2001; 2001US-0279036P.
 PR 30-MAR-2001; 2001US-0280233P.
 PR 02-APR-2001; 2001US-0280802P.
 PR 02-MAY-2001; 2001US-0288052P.
 PR 02-MAY-2001; 2001US-0288066P.
 PR 02-MAY-2001; 2001US-0288228P.
 PR 17-MAY-2001; 2001US-0291766P.
 PR 07-JUN-2001; 2001US-0296693P.
 PR 08-JUN-2001; 2001US-0296856P.
 PR 05-JUL-2001; 2001US-0303230P.
 PR 05-JUL-2001; 2001US-0303237P.
 PR 08-AUG-2001; 2001US-0310913P.
 PR 13-AUG-2001; 2001US-0311978P.
 PR 14-AUG-2001; 2001US-0312191P.
 PR 16-AUG-2001; 2001US-0312916P.
 PR 17-AUG-2001; 2001US-0313182P.
 PR 20-AUG-2001; 2001US-0313626P.
 PR 21-AUG-2001; 2001US-0314018P.
 PR 27-AUG-2001; 2001US-0315227P.
 PR 10-SEP-2001; 2001US-0318403P.
 PR 10-SEP-2001; 2001US-0318510P.
 PR 14-SEP-2001; 2001US-0322296P.
 PR 14-SEP-2001; 2001US-0322360P.
 PR 27-SEP-2001; 2001US-0325378P.
 PR 09-NOV-2001; 2001US-0332486P.
 PR 07-MAR-2002; 2002US-00094886.
 XX (CURA-) CURAGEN CORP.
 XX Kekuda R, Tchernev VT, Liu X, Spyttek KA, Patturajan M,
 PI Burtess CB, Vernet CAM, Li L, Gorman L, Malyankar UM, Boldog FL;
 PI Guo X, Shenoy SG, Radigaru M, Taupier RJ, Miller CE, Casman SJ;
 PI Pena CE, Gangoli EA, Gusev V, Smathson G, Zerfusen BD, Gerlach V;
 PI Pochart F, Fernandes ER, Shmukets RA, Rastelli L, Spaderna SK;
 PI Larochele WJ, Zhong M, Khramtsov NV, Voss EZ, Herrmann JL;
 XX WPI; 2003-058423/05.
 XX N-PSDB; ABX70491.
 XX
 XX NOVX polypeptides and polynucleotides, useful for treating a syndrome
 PT related to a human disease associated with the NOVX polypeptide e.g.,
 PT cancer.
 XX
 XX Claim 1; Page 223; 413pp; English.
 XX
 XX The present invention relates to the isolation of novel human
 XX polypeptides referred to as NOVX (NOV1-NOV44), variants of these

CC proteins, and the polynucleotide sequences encoding them. The NOVX
 CC proteins of the invention are G-protein coupled receptor. (GPCR) related
 CC proteins. The sequences of the invention are useful in the manufacture of
 CC a medicament for treating a syndrome related to a human disease
 CC associated with the polypeptides e.g. cancer. ABU52311-ABU52408 represent
 CC the human NOVX proteins of the invention
 XX
 SQ Sequence 1931 AA;

Query Match 23.6%; Score 901.5; DB 6; Length 1931;
 Best Local Similarity 32.2%; Pred. No. 2.5e-48;
 Matches 242; Conservative 93; Mismatches 266; Indels 151; Gaps 22;

QY 2 IKITFRPDSADGMLLYNGKRVKSPNTLANRQDFISFGLVGRPEFRFDAGSGMATIR 61
 Db 1282 LALEFRALFQGLLLYNGA-----RGKDFLALALLDGRVQLRFDFTGSGPAVLIT 1330
 QY 62 HPTPLALGHHTVTLRSLSLQGLIYVNGTSGKFKQGLDNEEYLVGGYDP---Y 118
 Db 1331 SAVPVEFGQHRLELSRHRRTLSVDGETPVLGESPGTDGLNLDLDFVGGVPEQAAA 1390
 QY 119 GAIPKAGLSGFGVRELRIOGEEIVFHDNL-----TAHGISHC--PTCRDRPCQNG 170
 Db 1391 VALERTFVAGLRGCLRLDANNRQI---ELGIGPAAATRGSGVGECHGHPCLPMPCHGG 1447
 QY 171 GQCHDSESSYVCVCFAGFTGSRCEHSQALHCHPEACGPDATCVARPDGRGTYTCRHLGR 230
 Db 1448 APCQNLEAGRFHQCPFRGPTCADEKS-PCQPNPCHGAAFCRVLPEG-GAQCCEPLGR 1505
 QY 231 SGLRCEGVTVTPSLSGAGSYLALPALTNTHELR-----LDVEFKPLAP 276
 Db 1506 EGTFCQ----TASGGDGGSPFLA-DFNGFSLHLRLGLHTFARDLGERKVALEVFVLRGP 1559
 QY 277 DGVLLFSGKSGVEFDVSLAMVGHLEFRYELGSLAVLRSAEPLALGRWHRVSAERLN 336
 Db 1560 SGLLYNGKTDGKDFVSLARDLREFRYDLGKAAVIRSRFPVTLGATWFRVSLERNG 1619
 QY 337 KQGLRVNGRPLVRSRCKSQGLMHTLLYLVGGVPSVPLSPATNMAHFRGCVGEVSV 396
 Db 1620 RKGALRVGDPRVLGSPVPHVTLNKEPLYVGGAPDFSKLARAASVSGFDGAIQLVSL 1679
 QY 397 NGKRL-----DLTYSFLSGSIGQCYDSSPCR---OPCQHGATCMPAGEYEFQC 443
 Db 1680 GGRQLLTPHVLQVDVT-SFAG-----HFCTRASGHPCLNGASVCP----- 1720
 QY 444 LCEDFGKGLCEHENPCQLRPECLHGTCQTRCLCPGFSGPRCQGGSGHGIABSDMW 503
 Db 1721 -----REAA-----VLCRFGSFGPCHCKGL----- 1742
 QY 504 LEGSGNDAPQGYGAYFDDGFLAPPVHFRSRLPEVPET-----LELEVRTSTASG 555
 Db 1743 VEKSAG-----DVTDLAFDGRFTVEYLNNAVTESEKALQSNHPELSLRTFATQG 1790
 QY 556 LLLWQVEVGEAGQKDFISLGLQDHLVFRYQLGSGEARLVSEDPINDGEMHVRTALRE 615
 Db 1791 LVLWS-----GKATERADYVALAIVDGHLSQLSNGLSQPWLNRSTVFNWNRWVAHRE 1846
 QY 616 GRGSIQVDEGELVGRSGPNNVAKGSYVIGGAPD--YATLTGGRFSPSGITGVKVL 673
 Db 1847 QREGSLQVNEAPVTVGSSPLGATQDLDGALWGLGPELVPVGPALPKAYGTGFGVGLRDV 1906
 QY 674 VLHSARPAPPQPLDLOHRAQAGANTRPFCS 705
 Db 1907 VVGR-----HPLHLEDAVTKPELRPCPT 1930

RESULT 6
 ID ADC39154
 AC ADC39154;
 XX
 DT 18-DEC-2003 (first entry)

XX Novel human NOVX polypeptide SEQ ID NO: 96.
 DE anti-diabetic; cytostatic; immunomodulator; anorectic; antilipemic;
 XX neurotropic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;
 KW antiasthmatic; antiinflammatory; hypotensive; antiarteriosclerotic;
 KW hemostatic; osteopathic; gene therapy.; NOVX; diabetes; obesity; cancer;
 KW lymphoma; uterus cancer; prostate cancer; dyslipidemia; anorexia;
 KW wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;
 KW cardiomyopathy; AIDS; asthma; Crohn's disease; multiple sclerosis;
 KW hypertension; atherosclerosis; hemophilia; graft-versus-host disease;
 KW Albright hereditary osteodystrophy.

OS Homo sapiens.
 XX WO2003010327-A2.
 XX 06-FEB-2003.
 XX 02-MAY-2002; 2002WO-US014199.
 PR 02-MAY-2001; 2001US-0288063P.
 PR 03-MAY-2001; 2001US-0288395P.
 PR 07-MAY-2001; 2001US-0289087P.
 PR 09-MAY-2001; 2001US-0289817P.
 PR 09-MAY-2001; 2001US-0289818P.
 PR 11-MAY-2001; 2001US-0290194P.
 PR 14-MAY-2001; 2001US-0290753P.
 PR 15-MAY-2001; 2001US-0291161P.
 PR 16-MAY-2001; 2001US-0291243P.
 PR 18-MAY-2001; 2001US-0292001P.
 PR 21-MAY-2001; 2001US-0292374P.
 PR 22-MAY-2001; 2001US-0292587P.
 PR 23-MAY-2001; 2001US-0293107P.
 PR 25-MAY-2001; 2001US-0293747P.
 PR 29-MAY-2001; 2001US-0294109P.
 PR 29-MAY-2001; 2001US-0294110P.
 PR 30-MAY-2001; 2001US-0294434P.
 PR 31-MAY-2001; 2001US-0294827P.
 PR 12-JUL-2001; 2001US-0304879P.
 PR 14-AUG-2001; 2001US-0308901P.
 PR 17-AUG-2001; 2001US-031270P.
 PR 17-AUG-2001; 2001US-0313416P.
 PR 10-SEP-2001; 2001US-0318463P.
 PR 27-SEP-2001; 2001US-0325683P.
 PR 18-OCT-2001; 2001US-0330292P.
 PR 28-NOV-2001; 2001US-0333873P.
 PR 03-DEC-2001; 2001US-0336909P.
 PR 03-DEC-2001; 2001US-0337552P.
 PR 21-FEB-2002; 2002US-0359245P.
 PR 01-MAY-2002; 2002US-00136826.

(CURA-) CURAGEN CORP.
 Muller CE, Kekuda R, Malyankar UM, Li L, Pena CEA, Spytek KA;
 Gorman L, Guo X, Fernandes ER, Smithson G, Stone DU, Zerhusen BD;
 Patturajan M, Anderson DW, Mezes PS, Feyman JA, Maccougall JR;
 Padigaru MR, Rastelli L, Shenoy SG, Gerlach VL, Shimkets RA, Zhong M;
 Edinger SR, Eilerman K;
 WPI; 2003-239445/23.
 N-PSDB; ADC39153.
 PT New NOVX polypeptides and polynucleotides, useful in gene therapy,
 particularly for treating or preventing a syndrome associated with a
 human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,
 hypertension or hemophilia.
 XX Claim 1; SEQ ID NO 96; 748pp; English.
 CC The invention relates to new isolated NOVX polypeptides, the genes
 CC encoding them or sequences having at least 95% identity to the amino acid
 CC or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,

CC particularly in the manufacture of a medicament for treating a syndrome
 CC associated with a human disease, which includes a pathology associated
 CC with NOVX polypeptide. The NOVX polypeptide is particularly useful for
 CC treating, preventing or alleviating pathology associated with NOVX
 CC polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and
 CC polypeptide are especially useful for treating or preventing e.g.
 CC diabetes, obesity, cancers (e.g. lymphoma, uterus cancer or prostate
 CC cancer), dyslipidemia, anorexia, wasting disorders, Alzheimer's disease,
 CC Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's
 CC disease, multiple sclerosis, hypertension, atherosclerosis, hemophilia,
 CC graft-versus-host disease or Albright hereditary osteodystrophy. The DNA
 CC encoding the protein is useful in gene therapy for treating the above
 CC conditions. These are also useful in developing powerful assay systems for
 CC functional analysis of various human disorders, as well as in diagnostic
 CC applications. This sequence represents one of the NOVX proteins of the
 CC invention.

XX Sequence 2053 AA;

Query Match 23.3%; Score 991.5; DB 7; Length 2053;

Best Local Similarity 32.1%; Pred. No. 1.2e-47;

Matches 243; Conservative 94; Mismatches 263; Indels 157; Gaps 24;

Qy	2	IKITFRPDSADGMLLYNGQKEVPGSPPTNLANRQPDFISFGVLCGRPFPRFDAGSGMATIR	61
Db	1400	LALFRALEPQGLSYNGNA-----RGKDFLALALLDGRVQLRFDFTSGSFAVLT	1448
Qy	62	HPTPLALGHFTVTLRLSLRTOGSLVIGDLAPVNGTSQKFGQLDLNEELYLGGYDP--Y	118
Db	1449	SAVPEPQWHERLSRHRRTLSVDCETPVLSGESPGTDLNLDLDFLVGGVPEPQAA	1508
Qy	119	GAIPKAGLSSGFICVRELRLQGEIEVFDHNL-----TAGIISHC--FTCDRRCQNG	170
Db	1509	VALERTFVAGLRICIRLLDNNORL---ELGIFGAATRSGVGECCDRPCLPNPCHGG	1565
Qy	171	GQHDSESSYVCPAGFTGSRCEHSQALHCHPEACCPDATCNRRDGRGYTCRCHLR	230
Db	1566	APQCNLEAGRPHCCPPGRVPTCADEKS-PCQENPCHGAAPCVLPEG-GAQCCEPLGR	1623
Qy	231	SLRCEBQVYVTTSLSAGSYLALPALTWTHHLR-----LDVEFKPLAP	276
Db	1624	EGTFQ-----TAGSQDGSQGFPLA-DFNGFSLHLEGLHTIARDLGEKMALEAVFLARGP	1677
Qy	277	DGVLFFSGKSGPVEDFVSLAMVGHLEFRYELGSLAVLRSAPLALGRWHRVSAERLN	336
Db	1678	SGLLLYNQKTDGKDFVSLARLDRLEFRYDLGKGAIVRSREPVTLGAWTRVSLERNG	1737
Qy	337	KDGLRVNGRPVLRSSFGKSGQ-----LMHHTLLYLGVEVPSVPLSPATNMSAHPRCV	391
Db	1738	RKCALRVGDGPRVLGESP-KSRKVPHTVNLNKEPLVYGGAPDFSKLARAASVSGFDGAI	1796
Qy	392	GEVSVNGKRL-----DLTVSFLGSGQIGQCYDSSPCER--QPCHGATCPAGE	438
Db	1797	QLVSLGSRQLLTPHVLNRQVDVT-SFAG-----HPCTRASGHFLNGHASCVP---	1842
Qy	439	YFOCLCRDGFKGLDCEBHENPCOLRPELCHGFTCOQTRCLLFGSGPRCQQGSGHGA	498
Db	1843	-----REAAV-----VCLCPGFSGPHCKEKL-----	1864
Qy	499	ESDWHLREGSGNDAPQGYAFHDDGFLAFPHVFRSRLPEVPEF-----IELEVT	550
Db	1865	-----VEKSAQ-----DVTDLAFDGRFTFVYMLNAVTESEKALQSNHFLSLRT	1907
Qy	551	STAGLLWQVEYGEAGQKDFISLGLQDGHVFRYQLGSGEARLVSDPINDGSEWHRV	610
Db	1908	EATQGLVWS---GKATERADVVALAIVDGHLQLSNTLNGSQPVLVLRSTVYVNTNRWLRV	1963
Qy	611	TALREGREGSTOYDGEELVSGRSPGNVAVNAKSGVYIGAPP--VATLTGGRFSSGIGT	668
Db	1964	VAHQREGSLQVNGEAPVYVSSPQATQDLDGALMLGLPGLPVPALPKRAYGTGVFG	2023
Qy	669	CVKNLVLSARPAGPPPPQPLDQHRQAQAGANTSPFCS	705

Db 2024 CLRDDVVGR-----HPLLHLEDAVTKPELRPCPT 2052

RESULT 7

ADC39164

ID ADC39164 standard; protein; 2143 AA.

AC ADC39164;

XX AC

DT 18-DEC-2003 (first entry)

XX

Novel human NOVX polypeptide SEQ ID NO: 106.

XX

anti-diabetic; cytostatic; immunomodulator; anorectic; antilipemic;

XX nootropic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;

KW antiasthmatic; antiinflammatory; hypotensive; antiarteriosclerotic;

KW hemostatic; osteopathic; gene therapy; NOVA; diabetes; obesity; cancer;

KW lymphoma; uterus cancer; prostate cancer; dyslipidemia; anorexia;

KW wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;

KW cardiomyopathy; AIDS; asthma; Crohn's disease; multiple sclerosis;

KW hypertension; atherosclerosis; hemophilia; graft-versus-host disease;

KW Albright hereditary osteodystrophy.

XX

OS Homo sapiens.

XX

WC02003010327-A2.

XX

PD 06-FEB-2003.

XX

PF 02-MAY-2002; 2002NO-US014199.

XX

PR 02-MAY-2001; 2001US-0288063P.

PR 03-MAY-2001; 2001US-0288395P.

PR 07-MAY-2001; 2001US-0289087P.

PR 09-MAY-2001; 2001US-0289817P.

PR 09-MAY-2001; 2001US-0289818P.

PR 11-MAY-2001; 2001US-0290194P.

PR 14-MAY-2001; 2001US-0290753P.

PR 15-MAY-2001; 2001US-0291181P.

PR 16-MAY-2001; 2001US-0291243P.

PR 18-MAY-2001; 2001US-0292001P.

PR 21-MAY-2001; 2001US-0292374P.

PR 22-MAY-2001; 2001US-0292587P.

PR 23-MAY-2001; 2001US-0293107P.

PR 25-MAY-2001; 2001US-0293747P.

PR 29-MAY-2001; 2001US-0294109P.

PR 29-MAY-2001; 2001US-0294110P.

PR 30-MAY-2001; 2001US-0294434P.

PR 12-JUL-2001; 2001US-0294827P.

PR 31-JUL-2001; 2001US-0304879P.

PR 14-AUG-2001; 2001US-0312270P.

PR 17-AUG-2001; 2001US-0313416P.

PR 10-SEP-2001; 2001US-0318463P.

PR 27-SEP-2001; 2001US-0325683P.

PR 18-OCT-2001; 2001US-0330292P.

PR 28-NOV-2001; 2001US-0333873P.

PR 03-DEC-2001; 2001US-0336909P.

PR 03-DEC-2001; 2001US-0337552P.

PR 21-FEB-2002; 2002US-0359245P.

PR 01-MAY-2002; 2002US-00136826.

XX

(CURA-) CURAGEN CORE.

PA

XX Miller CE, Kekuda R, Malyankar UM, Li L, Pena CE, Spyttek KA;

XX Gorman L, Guo X, Fernandes ER, Smithson G, Stone DJ, Zerrhusen BD;

PI Paccurajan M, Anderson DW, Mezes FS, Peyman JA, Maccougall JR;

PI Padigaru M, Rastelli L, Shenoy SG, Gerlach VL, Shimkets RA, Zhong M;

PI Edinger SR, Ellerman K;

XX WPI; 2003-239445/23.

DR N-PSDB; ADC39163.

XX

CC The present sequence is the protein sequence of Incyte polypeptide
 CC 2736276CD1 denoted human cell adhesion and extracellular matrix protein 8
 CC (CADECM-8). The protein is encoded by a clone isolated from a soft tissue
 CC tumour cDNA library. The invention provides CADECM-1 to -11 polypeptides
 CC (see ABP58224-34) and polynucleotides (see ABZ24578-88), vectors, host
 CC cells, antibodies, agonists and antagonists. These are useful for
 CC diagnosing, treating or preventing disorders associated with aberrant
 CC expression of CADECM, particularly cell proliferative disorders (e.g.
 CC osteoarthritis, atherosclerosis, cirrhosis, hepatitis, paroxysmal
 CC nocturnal haemoglobinuria, polycythaemia vera, psoriasis, primary
 CC thrombocytopenia or cancer), developmental disorders (e.g. renal
 CC tubular acidosis, anaemia or mental retardation), neurological disorders
 CC (e.g. Alzheimer's disease, Parkinson's disease or epilepsy), reproductive
 CC disorders (e.g. infertility or a disruption in the menstrual cycle), or
 CC autoimmune/inflammatory disorders (e.g. AIDS, allergy, asthma, autoimmune
 CC thyroiditis, contact dermatitis, Crohn's disease, diabetes mellitus,
 CC Glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease,
 CC Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis,
 CC osteoarthritis, osteoporosis, pancreatitis, Reiter's syndrome, rheumatoid
 CC arthritis, Sjogren's syndrome, uveitis) or viral, bacterial, fungal,
 CC parasitic, protozoal or helminthic infections
 XX
 SQ Sequence 671 AA;

```

Query Match      22.0%; Score 840; DB 6; Length 671;
Best Local Similarity 29.6%; Pred. No. 6e-45;
Matches 214; Conservative 107; Mismatches 257; Indels 144; Gaps 20;

QY 1 EIKITFRDSDADGMLLYNQGKRVGSPNLANRQPDIISFGLYGGPEREFDAGSGMWTI 60
Db 1  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 72 QITLFRAEADGMLLYCGENE-----HGRGDFMSLAIIRSLQFRFCGTVGVAII 122
Db 72  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 61 RHPTPLALGHPHTVLRSLLTQSSLIYVGLDPLAIVNGTSGQKFGQLDLNEELYLGGYPD-YG 119
Db 61  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 123 VSETKILGGWHYVLRDGLNGLQLNGTVPVTSQSGQYVKITRTPLYLGGAPSAYW 182
Db 123  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 120 AIPKAGLSSGRTGVRELRIOGEBRIVFHDLN-----LTAHGISHCPT--CRDRCOQG 170
Db 120  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 183 LVATGTRNGFQGVQSLVAVNGRRI---DWRPFLGKALSGADVGESSGICDEASCIWG 239
Db 183  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 171 GOCHDSESSVYVCPAGTSGRCSEHSQALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
Db 171  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 240 GTCTAKADSVICLPLGF----- 258
Db 240  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 231 SLRCEEGYVTTPRSLSGAGSYLALPALTWTHHEL--RLDVEFKPLAPDGVLLFS--- 283
Db 231  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 259 KGRHCEDAFTLIPOFRSLRSYATWPPELPQHYLSFMFEFTFRPDSGDGLLYSYDT 318
Db 259  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 284 GKSQGFVEDFVSLAMVGGHLEFRYELGSLAVLRSAPALGRWHRVSAERLNKDGSLRV 343
Db 284  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 319 GSK-----DFLSINLAGHVBFYDFDCGSGTGLRSAPLPLTIGNWHELVRVSTAKNGI 373
Db 319  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 344 NGGRLPVRSSPKSQQGLNLHLLYLGVEPSPVLSPATNMSAHRGCVGVSVNGKRLDL 403
Db 344  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 374 DKQKIVEGMAEGGFTQIKNTDIFIGVFNVDYVKNSSVLPKPSGIQKIIINDRTHIV 433
Db 374  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 404 TVSFYLSGGIQCYSPPCRQPOHGATCMPAGEYEFQCLCRDGFGKDLCEHEENPCQL 463
Db 404  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 434 KHDF--TSGVNVENAAHPCARVCAHGGSCRPKE-GYVDCDCLPFGF----- 478
Db 434  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 464 REPCLHGTCQGT--RCI-CLGFGSPRCQSGSHGIAESDHWLBSGSGNDAPGCGAYFH 521
Db 464  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 479 ----LH---CQXATEAEIEIYQFIQ-----RSYLYTDN 504
Db 479  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 522 DGFLLAFPHGFVRSLPVEPTEILEVETSTAGSLLLWQGVGEAQQKDFISLGLQDG 581
Db 522  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 505 PDILKRVSSG--SRSS-----NVFMRFKTKADGILLWRG--DSPMFPNSDFISLGLR 553
Db 505  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 582 HLVFVYQLGGSEARLVSDPNDGSHWRVVTALRGRGSIQVDGEEELVSRSPGNVAVN 641
Db 582  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 554 ALVFSYNLGGVASTWVNGSFDNDRWHRVAVRQGGSGKITVDDYGARTGKTSFGMWRQLN 613
Db 554  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 642 AKGSVIYGAPDVAITLGRSSGSGITGCVXNVLNLSRPARPPQPLDLQHRQAQGNTR 701
Db 642  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

```

Db 614 INGALVYGGKMEIAFHNRQVMEGLVGCISHFTLSI-----DYHISLVDEAVDGKNIN 666
QY 702 FC 703
Db 667 TC 668

RESULT 9
ADE28105
ID ADE28105 standard; protein; 1009 AA.
AC ADE28105;
XX
XX 29-JAN-2004 (first entry)
XX Human NTRAN protein - SEQ ID 10.
XX human; neurotransmission-associated protein; NTRAN; cytostatic; immunomodulator; immune disorder; cancer; gene therapy.
XX Homo sapiens.
XX WO2003051902-A1.
XX 26-JUN-2003.
XX 12-DEC-2002; 2002WO-USO40059.
XX 14-DEC-2001; 2001US-0340798P.
XX 18-MAR-2002; 2002US-0365645P.
XX 25-MAR-2002; 2002US-0367862P.
XX 10-MAY-2002; 2002US-0379887P.
XX 31-MAY-2002; 2002US-0384639P.
XX (INCY-) INCYTE GENOMICS INC.
XX Baughn MR, Bhatia U, Blake JJ, Burrill JD, Elliott VS, Emrling BM, Forsythe IJ, Gietzen KJ, Gorvad AB, Griffin JA, Hafalila AUA, Ho A, Jackson AA, Jiang X, Kabie AB, Kearney L, Khare R, Lee EA, Lee S, Lu DAM, Marquis JP, Lehr-Mason PM, Rankumar J, Richardson TW, Sprague WW, Tran UK, Chawla NK, Warren BA, Yue H, Zheng W;
XX WPI; 2003-514037/48.
XX N-PSDB; ADE28127.
XX
XX New human neurotransmission-associated proteins (NTRAN) polypeptide, useful for preparing a composition for treating a disease associated with decreased expression or overexpression of NTRAN e.g., cancer.
XX Claim 1; SEQ ID NO 10; 261pp; English.
XX
XX The invention relates to a novel isolated human neurotransmission-associated proteins (NTRAN) polypeptide. The polypeptide of the invention demonstrates cytostatic and immunomodulator activities and may be useful for preparing a composition for diagnosing or treating a disease or condition associated with decreased expression or overexpression of functional NTRAN including immune disorders or cancer, as well as during gene therapy procedures. The current sequence is that of the human NTRAN protein of the invention.
XX Sequence 1009 AA;

```

```

Query Match      22.0%; Score 840; DB 7; Length 1009;
Best Local Similarity 29.6%; Pred. No. 9.6e-45;
Matches 214; Conservative 107; Mismatches 257; Indels 144; Gaps 20;

QY 1 EIKITFRDSDADGMLLYNQGKRVGSPNLANRQPDIISFGLYGGPEREFDAGSGMWTI 60
Db 410 QITLFRAEADGMLLYCGENE-----HGRGDFMSLAIIRSLQFRFCGTVGVAII 460
QY 61 RHPTPLALGHPHTVLRSLLTQSSLIYVGLDPLAIVNGTSGQKFGQLDLNEELYLGGYPD-YG 119

```


Db 461 VSETKIKLGGWHTVLMYLDGLNGLLNNGTPTVGGQGOYSKIIFRTPLELYLGGPAPAYW 520
 QY 120 AIPKAGLSSFGICVRELRIOGEBEIVFHDLN-----LTAHGISHCPT--CRDRPQNG 170
 Db 521 LVRATGTRNGFOCVQSLAVNGRRI---DMRPWFLKALSGADVGECSGICDEASCING 577
 QY 171 GQCHDSESSVYCVCPAGFTSRCEHSOALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
 Db 578 GTCATKADSYICLCPGFG----- 596
 QY 231 SGLRCEEGVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFS--- 283
 Db 597 KGRHCEDAFTLTIPOPRESLSRYAATPWLPFPQHYLSFMFEFETFRPDSGDVLLYSYDT 656
 QY 284 GKSGVEDFVSLAMVGGHLEFRYELGSLAVLSAEPALGRWHRVSAERLNKDGSLRV 343
 Db 657 GSK-----DFLSINLAGHVEFRFCGSGTGLVRSDEPLTLGNWHELAVSRKNGILQV 711
 QY 344 NGRPVLRSPPKSOGLNHLTLLYLGGVPSVPLSPATNMSAHFRGCVGEVSNKRLDL 403
 Db 712 DKQKIVGEMAGGGFTQIKNCNTDIFIGVFNVDVKKNSGVLKPFSSIQKIILNDRITHV 771
 QY 404 TVSYFLSGOIGQCYDSSPCERQPCQHGATCMPAGEYEFQCLCRDGFKDLCEHBEPCQL 463
 Db 772 KHDF--TSGVNVENAAHPCVRAPCAHGSGCRPRKE-GYDCDCPLGFG----- 816
 QY 464 REPCLHGTCQGT--RCLCLPGFSPRCQSGHGIASDWHLESGGNDAPGOYAYFH 521
 Db 817 ----LH---CQKAIETAEIIEIPIQFIG-----RSYLVYDN 842
 QY 522 DDGFLAFPGHVFHSRSLPEVPEFTELEVRTSTASGLLLMQGVEGVEAGCKDFISLGLQDG 581
 Db 843 PDILKRVSG---SRS-----NVFMRFTAKDGLLLWRG--DSPMRPNSDFISLGLRDG 891
 QY 582 HLVFRYQLGSGEARLVSEDPINDGEMWHRVTLALREGRGSIQVDEBELVSGRSPGNVAVN 641
 Db 892 ALVFSYNLGGVASTIMVNGSFNDRGHRVKAVRDQSGKIVDDYDYGARTGKSPGMWQLN 951
 QY 642 AKGSVYIGCAPDVAFLTGTGFSSTGTCVKNLVLSARAPGAPPQPLDQHRQAQANTR 701
 Db 952 INGALYVGMKEIALHTNRQWRGLVGCISHFTLST-----DYHISLVEDAVDGRKIN 1004
 QY 702 PC 703
 Db 1005 TC 1006
 RESULT 10
 ABB72291
 ID ABB72291 standard; protein; 819 AA.
 AC ABB72291;
 XX ABB72291;
 DT 04-APR-2002 (first entry)
 XX
 DE Rat protein isolated from skin cells SEQ ID NO: 503.
 XX
 KW Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
 XX developmental defect; inflammatory disease; dermatological; vulnerary;
 XX immunomodulator; anti-inflammatory; cytostatic; neuroprotective.
 OS Rattus sp.
 XX
 EN W0200190357-A1.
 XX
 PD 29-NOV-2001.
 XX
 PF 24-MAY-2001; 2001WO-NZ000099.
 XX
 PR 24-MAY-2000; 2000US-0206550P.
 PR 25-JUL-2000; 2000US-0221232P.
 XX

(GENE-) GENESIS RES & DEV CORP LTD.
 Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG,
 Kumble KD,
 WPI; 2002-122020/16.
 New polynucleotides and polypeptides encoded by the polynucleotides
 isolated from skin cells, useful for treating skin wounds, cancers,
 growth and developmental defects, inflammatory diseases, or for
 modulating immune responses.
 Claim 4; Page 308-310; 466pp; English.
 The present invention provides the protein and coding sequences of cDNAs
 isolated from human, murine and rat skin cell libraries. The sequences
 can be used in the development of therapeutic agents useful in the
 treatment of skin diseases, including skin wounds, cancer, growth
 defects, developmental defects and inflammatory diseases. The proteins
 have important roles in the induction of hair growth, cell proliferation
 and cell-cell interaction, in maintaining tissue integrity, in wound
 healing and in modulating immune responses. The present sequence is a
 polypeptide of the invention
 Sequence 819 AA;
 Query Match 21.8%; Score 832; DB 5; Length 819;
 Best Local Similarity 29.3%; Pred. No. 2.4e-44;
 Matches 210; Conservative 109; Mismatches 264; Indels 134; Gaps 17;
 QY 1 EIKITRPSADGMLLYNQKRVPSPTNLNRQDFDFISGLVGRPFRFAGSGMAII 60
 Db 220 QITLEFRAEADGLLLYCGESE-----HGRGDEMSLALIRRSLHFRFCGIGMAII 270
 QY 61 RHPTLALGHEHTVTLRSITOGSLIVGLAVNGTSQKFOGLDNLNEELYLGGYDP-YG 119
 Db 271 ISETKLGARHSVLYRDGLNGLQLANGTPTVGGQGOYSKIIFRTPLELYLGGPAPAYW 330
 QY 120 AIPKAGLSSFGICVRELRIOGEBEIVFHDLN-----LTAHGISHCPT--CRDRPQNG 170
 Db 331 LVRATGTRNGFOCVQSLAVNGRRI---DMRPWFLKALSGADVGECSGICDEASCING 387
 QY 171 GQCHDSESSVYCVCPAGFTSRCEHSOALHCHPEACGPDATCVNRPDGRGYTCRCHLGR 230
 Db 388 GTCATKADSYICLCPGFG----- 409
 QY 231 SGLRCEEGVTTPSL-SGAGSYLALPALTNTHHEL---RLDVEFKPLAPDGVLLFSGCK 286
 Db 410 ----HCEDAFTLTIPOPRESLSRYAATPWLPFPQHYLSFMFEFETFRPDSGDVLLYSYDT 466
 QY 287 SGPVEDFVSLAMVGGHLEFRYELGSLAVLSAEPALGRWHRVSAERLNKDGSLRVNGG 346
 Db 467 SS--KDFLSIIAAGHVEFRFCGSGTGLVRSDEPLTLGNWHLAVSRKNGILQV 524
 QY 347 RPYLRSPPKSOGLNHLTLLYLGGVPSVPLSPATNMSAHFRGCVGEVSNKRLDLTYS 406
 Db 525 KVEGMAEGGFTQIKNCNTDIFIGVFNVDVKKNSGILHFPSSGIQKLIILNDRITHV 584
 QY 407 FLGSGICQCVDSFPCRPQCHGATCMPAGEYEFQCLCRDGFKDLCEHBEPCQLREP 466
 Db 585 F--TSGVNVENAAHPCVRAPCAHGSGCRPRKE-GYDCDCPLGFGELNCKAI----- 633
 QY 467 CLHGTCCQTRCLCLPGFSPRCQSGHGIASDWHLESGGNDAPGOYAYFHDDGFL 526
 Db 634 -----TEAIEIPQIGSYLYTNPMLK---RVSGRSN-----AFM 668
 QY 527 APGHVFSRSLPEVPEFTELEVRTSTASGLLLMQGVEGVEAGCKDFISLGLQDGHVIFVR 586
 Db 669 RF-----KTTAKDGLLLWRG--DSPMRPNSDFISLGLRDGALVFS 706
 QY 587 YOLGSGEARLVSEDPINDGEMWHRVTLALREGRGSIQVDEBELVSGRSPGNVAVNAGSV 646
 Db 707 YNLGSGVASIMVNGSFNDRGHRVKAVRDQSGKIVDDYDYGARTGKSPGMWQLNIN 766

```

QY 647 YTGADPVATLGGFSSGIGTCVKNLHLSARPGAPPPDLDLQHRQAQANTRPC 703
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 767 YVGMKEIALHNRQYMRGLVGCISHFTLSI-----DYHISLVEDAVGKNVNTC 816
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
RESULT 11
ABB63614
AC ABB63614 standard; protein; 4072 AA.
XX
XX
XX 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 17634.
XX
KW KX
KW KX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US009231.
PF
XX 23-MAR-2000; 2000US-0191637P.
PR
XX 11-JUL-2000; 2000US-00614150.
XX
FA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL07717.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
PS Disclosure; SEQ ID NO 17634; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB5773-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4072 AA;
Query Match 21.2%; Score 812.5; DB 4; Length 4072;
Best Local Similarity 30.2%; Pred. No. 2.6e-42;
Matches 219; Conservative 121; Mismatches 304; Indels 81; Gaps 22;
QY 3 KITFRPDSADGMLLYNGOKKVPQSPNLANRPDRISFGLVGRPEFRFDAGSGMATIRH 62
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 3394 ELTFRENGDGLLFWQTRGSG-----DYLALSLEKRYAEFRFDGKPLMIVRA 3443
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 63 PTLALGHFTVTLNLSLTCGSLVIGDLAPVNGTSGQKFGQLDNBELYLGGYPDYGAIP 122
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 3444 EEPFALNEWHTVRSVFRKRDYIQVDEQHPVAFPTLQQIPQLDLIEDLYIGGVNPELLP 3503
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 123 KAGLSS--GFIGVRLRIGGEEIVPHDLNLTAHGISHCPTCEDRCPQGGCHDSES-S 179
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 3504 ADAVSOQGFVGCISRLTLQRTVELIREAKYKEGITDCRCAQGPCQNKVCLESQTEQ 3563
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 180 SYVCVCPAGFTGRCSBHSQALHCHPEACGPDATCNVRPDGRGTYTCRHLGRSGLRCBEV 239
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:

```

```

Db 3564 AITCICQPWTRDCA-IEGTQCTFVCGA-GRCENTEN--DMECLCFNRSRGCRCQYNE 3619
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 240 TVTTPSLSGAGSYLA---LPALTNTHHELRLDVEEKPLA-PDGVLLFSGGKGPVEDFVY 295
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 3620 ILNHSLSNFKGNSFAAYGTFKVT---KVNITLSVVRPASLESDVILYTAESTLPSGDYLA 3675
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 296 LAMVGGHLEFRYELGSL--AVLRSABPLALGRWHRVSAERLNKDGSLRVNGRRPVLRS 353
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 3676 LVIRGGHALLLNTAARLDVAVRVAEPLPLNRWTRIEIRRRRLRGGELRVGDPGRKAKA 3735
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 354 PGRSQGLNLHLLYLGVEPS-VPLSPATNNSAHPRCGCVGEVSVNGKELDLTYSFLSGO 412
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 3736 PGSDRILLSLTKHLYVGGYDRSVKVRDVTNITKGFDDGCSRLYNYFKQKPNLLADIKDA 3795
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 413 IQCYDSSPCERQPCQHGATCMPAGEYE-----FQCLCRDGFKGDLCHEHENPCQLREP 466
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 3796 IQSC-----GETNMGDESDNEPVPVPTPDVHENELQFYAMAPC-ASDP 3841
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 467 CLHGGTCQGTR----CLCLPFGSQRCCQGGHGIAESDWHLEGGSGNDAPCQYGVAFHD 522
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 3842 CENGGSCSEQEDVAVCSQFPFGSGKHCB-----HLQLG-----FNASFRG 3882
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 523 DGFAPPGHVRSRSLPEVPEFIELEVRTSTASGLLMOGVVGRAGQKDFISLGLDGH 582
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 3883 DGYVELNRSHPALEQSYTSMGIVFTTKENGLLFWMQBAGEEYTCQDFIAAAVVDGY 3942
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 583 LVFRYQLGSGEARLVSED-PINDGEWHRVTALRBRGRSGIOVDGEEELVSGRS-PGPVAV 640
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 3943 VEYSNRLDGEAEVIRNSDIRVDNGERHIVIAKRDENTAILLEVD-RMLHSGEFRTRTSKSM 4001
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 641 NAKGSVYIGGAPDVATLGGFRFSSGITGCVKXNVLHLSARPGAPPQLDLOHRAQAGANT 700
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 4002 KLPQNVFVGGAPDLEVFTRFYRKHNLNGCI--VVVEGETVG-----QINLSSAAAVGVNA 4054
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 701 RPCPS 705
   ||:|:
Db 4055 NVCPA 4059
   ||:|:
RESULT 12
ABB61948
ID ABB61948 standard; protein; 1298 AA.
AC ABB61948;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 12636.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX N-PSDB; ABL06051.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX

```


DR N-PSDB; AAT90471.
 PT Nucleotide sequences encoding human agrin and muscle specific kinase and related receptor - used in diagnosis and treatment of disorder with muscle atrophy.
 XX
 PS Claim 2; Fig 15; 120pp; English.
 CC This polypeptide comprises the amino acid sequence of human agrin deduced from an isolated foetal brain cDNA clone (see AAT90471). Full-length agrin, or active portions of it, can be expressed using a claimed host-vector system. The agrin polypeptides can be used in claimed methods of treating diseases or disorders that affect muscle, especially atrophy resulting from denervation due to nerve trauma, degenerative, metabolic or inflammatory neuropathy, peripheral neuropathy or damage to nerves caused by environmental toxins or drugs. In particular the muscle atrophy is due to motor neuropathy, chronic disease, metabolic stress or nutritional insufficiency, muscular dystrophy syndrome, congenital myopathy or acquired (toxic or inflammatory) myopathy. The polypeptides can also be used in a claimed method of promoting the growth, differentiation or survival of cells that express muscle-specific (MuSK) receptors (see AAW26610-11), and to raise antibodies

Qy 202 CHPEACGPDACVNRPDGRGVTCTCHLGRSLRCEEGVTVTPSLGAGSVLALPALNT 261
 Db 27 CQNPFCGAAFCRVLPFG- GAQCEPLQREGTFQC-----TASQDGGFLA-DFNQFS 79
 Qy 262 FHELRL-----LDVEFKPLAPDGVLLFSGKSGFVDFVSLAMVGHLEFFRY 307
 Db 80 HLELRGLHTFARDLGERKMALEVFELARGPSGLLYNGOKTDGKDFVSLALDRRLREFFRY 139
 Qy 308 ELGSLAVLRGAELALGHRVSAERLNKDGSRVNGRVRPLRSSPKSQG-----LNL 362
 Db 140 DLGKGAIVIRREPVTILGAWTRVSLERNRKGALRVGDGPFVLSGSP-KSRKVPHTVNL 198
 Qy 363 HTLLYLVGVEFSVPLSPATNMSAHFRGCVGVSVNGKRL-----DLTVSFLGSGQ 412
 Db 199 KEPLVVGAPPDFSKLARAASVSSGFDGAIQLVSLGROLLTPEHVLVRQVDVT-SFAG--- 254
 Qy 413 IGQCVDSPFER---QPCHGATCMPAGEYEFQCLCRDGFKGDLCHEENPCQLREFLH 469
 Db 255 -----HPCTRASGHPCINGASCVP-----REAAV- 278
 Qy 470 GGTCQGTCLCLLPFGSPRCQGGSHGIAESDWHLEGGNGNDAPGQYGFHDDGFLAFP 529
 Db 279 -----VCLCPGSPGPHCEKGL-----VEKSAG-----DVDTLAFD 309
 Qy 530 GHVFRSLPEVPET-----LELVRFSTASGLLWQGVVEGAGQKDF 573
 Db 310 GRTFVEYLNVAVTESELANEIIPVEKALQSNHPELSLRTEATQGLVLS-----GKATEADY 365
 Qy 574 ISLGLQDGHVFRYQLGSGEARLVSEEDINDGEHVRVTLRREGRRGSIQVDGELVSGRS 633
 Db 366 VALAIVDGHQLSINLVSQPVVLRVTPVNTNMLRVVAHREQRGSLQVGNAPVIGSS 425
 Qy 634 PGPVNAVAKGSVYIGGAPD--VAFLTGGFRSSGHTGCKNVLVLSARAPGAPPQFDLQ 691
 Db 426 PLGATQLDIDGALWGLGLPELFPVGPALPKAYGTGFVGLRDVVGR-----HPLHLL 477
 Qy 692 HRAQAGANTRPCPS 705
 Db 478 EDAVTKPELRPCPT 491

Query Match 16.7%; Score 640.5; DB 2; Length 492;
 Best Local Similarity 33.0%; Pred. No. 1.9e-32;
 Matches 183; Conservative 59; Mismatches 173; Indels 139; Gaps 18;

Sequence 492 AA;

ADC39166;
 18-DEC-2003 (first entry)
 Novel human NOVX polypeptide SEQ ID NO: 108.
 anti-diabetic; cytostatic; immunomodulator; anorectic; antilipemic;
 nootropic; neuroprotective; immunostimulant; antiparkinsonian; anti-HIV;
 antiasthmatic; anti-inflammatory; hypotensive; antiarteriosclerotic;
 hemostatic; osteopathic; gene therapy; NOVX; diabetes; obesity; cancer;
 lymphoma; uterus cancer; prostate cancer; dyslipidemia; anorexia;
 wasting disorder; Alzheimer's disease; Parkinson's disorder; cachexia;
 cardiomyopathy; AIDS; asthma; Crohn's disease; multiple sclerosis;
 hypertension; atherosclerosis; hemophilia; graft-versus-host disease;
 Albright hereditary osteodystrophy.
 Homo sapiens.
 WO2003010327-A2.
 06-FEB-2003.
 02-MAY-2002; 2002WO-US014199.
 02-MAY-2001; 2001US-0288063P.
 03-MAY-2001; 2001US-0288395P.
 09-MAY-2001; 2001US-0289087P.
 09-MAY-2001; 2001US-0289817P.
 11-MAY-2001; 2001US-0289818P.
 14-MAY-2001; 2001US-0290194P.
 15-MAY-2001; 2001US-0291181P.
 16-MAY-2001; 2001US-0291243P.
 18-MAY-2001; 2001US-0292001P.
 21-MAY-2001; 2001US-0292374P.
 22-MAY-2001; 2001US-0292587P.
 23-MAY-2001; 2001US-0293107P.
 25-MAY-2001; 2001US-0293747P.
 29-MAY-2001; 2001US-0294109P.
 29-MAY-2001; 2001US-0294110P.
 30-MAY-2001; 2001US-0294434P.
 31-MAY-2001; 2001US-0294827P.
 12-JUL-2001; 2001US-0304879P.
 31-JUL-2001; 2001US-0308901P.
 14-AUG-2001; 2001US-0312270P.
 17-AUG-2001; 2001US-0313416P.
 10-SEP-2001; 2001US-0318463P.
 27-SEP-2001; 2001US-0325683P.
 18-OCT-2001; 2001US-0330292P.
 28-NOV-2001; 2001US-0333873P.
 03-DEC-2001; 2001US-0336909P.
 03-DEC-2001; 2001US-0337552P.
 21-FEB-2002; 2002US-0359245P.
 01-MAY-2002; 2002US-0013682E.
 (CURA) CURAGEN CORP.
 Miller CE, Kekuda R, Malyankar UM, Li L, Pena CEA, Spytek KA;
 Gorman L, Guo X, Fernandes ER, Smithson G, Stone DJ, Zerhusen BD;
 Pattarajan M, Anderson DW, Mezes PS, Peyman JA, Macdougall JR;
 Padigaru M, Rastelli L, Shenoy SG, Gerlach VL, Shinkets RA, Zhong M;
 Edinger SR, Ellerman K;
 WPI; 2003-239445/23.
 N-PSDB; ADC39166.
 New NOVX polypeptides and polynucleotides, useful in gene therapy,
 particularly for treating or preventing a syndrome associated with a
 human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,
 hypertension or hemophilia.
 Claim 1; SEQ ID NO 108; 748pp; English.

RESULT 14
 ADC39166
 ID ADC39166 standard; protein; 1566 AA.

XX The invention relates to new isolated NOVX polypeptides, the genes
 CC encoding them or sequences having at least 95% identity to the amino acid
 CC or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,
 CC particularly in the manufacture of a medicament for treating a syndrome
 CC associated with a human disease, which includes a pathology associated
 CC with NOVX polypeptide. The NOVX polypeptide is particularly useful for
 CC treating, preventing or alleviating pathology associated with NOVX
 CC polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and
 CC polypeptide are especially useful for treating or preventing e.g.
 CC diabetes, obesity, cancers (e.g. lymphoma, uterine cancer or prostate
 CC cancer), dyslipidemias, anorexia, wasting disorders, Alzheimer's disease,
 CC Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's
 CC disease, multiple sclerosis, hypertension, atherosclerosis, hemophilia,
 CC graft-versus-host disease or bright hereditary osteodystrophy. The DNA
 CC encoding the protein is useful in gene therapy for treating the above
 CC conditions. These are also useful in developing powerful assay system for
 CC functional analysis of various human disorders, as well as in diagnostic
 CC applications. This sequence represents one of the NOVX proteins of the
 CC invention.

XX SQ Sequence 1566 AA;
 Query Match 16.6%; Score 635.5; DB 7; Length 1566;
 Best Local Similarity 33.4%; Pred. No. 1.5e-31;
 Matches 179; Conservative 58; Mismatches 174; Indels 125; Gaps 16;

QY	207	CGPATCVNRDPGRGTYCRCHLGRSGLRCEGEVTTTPSLSGAGSYLALPALTWHHEIR	266
Db	1118	CHGAAPCRVLEPG-GAQCCEPLGREGTFQC-----TASGQDGGPFLLA-DFNGFSLHLELR	1170
QY	267	-----LDVEFKLADPGLVLLFSGHSGVPEFVSLAMVGGHLEFRYELGSG	312
Db	1171	GLHTFARDLGEKMALEVFLARGPSGULLYNGQTKDCKGDFVSLARDRLFRYDLGKG	1230
QY	313	LAVRSAPLALGRWHRVSRBLNKDGLRNGRPNVLRSPKSGQLNHLTLILGQVE	372
Db	1231	AAVIRSRPVTLGAWTRVSLRNGKRALRYGDGPRVLGSPVPHVTLNLKPELYVGGAP	1290
QY	373	PSVPLSPATNSAHPFGCVGVSVNGKRL-----DLTYSFLSGIGTCQYDSSPC	422
Db	1291	DFSKLARAASVSGFDGAIQLVSLGGRLTPEHVLQVDTV-SFAG-----HPC	1339
QY	423	ER---QPCGHGATCMPAGEYEFQCLCRDGFKGDLCHEHENPCQLRPECLHGTCQGRCL	479
Db	1340	TRASHGPLCNGASCVP-----REAAV-----VCL	1363
QY	480	CLPGFSGPRCOQSGHGHTAESDWHLEGGGNDAPGQYGAYFTHDDGFLAFPHVFSRSLPE	539
Db	1364	CPGFGSPHCEKGL-----VERSG-----VDVTLAFDGRTFVEYLNA	1401
QY	540	VPET-----IELEVRTSTAGLLMQVGEAGQKDFISLGLQDGHVFRYQLGS	591
Db	1402	VTESEKALQSNHFELSLRTEATQGLVLS-----GKATERADYVALAIVDGHQLSYNLGS	1457
QY	592	GEARLVSEDPINDEWHRVTALREGRGSIQVDGSELVSGSPFPVAVNAKGSVYIGGA	651
Db	1458	QPVVLRSTVPTVNRVLRVVAHREQREGSLQVGNAPVTVSSPLGATQDLDGALWIGGL	1517
QY	652	PD--VATLTGGRFSSGITGVKNLVLSARFGAPPFPQDLDLQHRQAQAGANTRPCPS	705
Db	1518	PELPGVPALPRAYGTGFGVCLRDVVVGR-----HPLHLEDAVTKPELRPCPT	1565
RESULT 15			
ADB64904			
ID	ADB64904	standard; protein; 463 AA.	
XX			
AC	ADB64904;		
XX			
DT	04-DEC-2003	(first entry)	
XX			
DE		Human protein encoded by clone PLAC860113340.	

XX Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
 KW cell regeneration; membrane protein; signal transduction-related protein;
 KW transcription-related protein; osteoporosis; neurological disease;
 KW cancer; tumour.
 XX Homo sapiens.
 PN EF1308459-A2.
 XX 07-MAY-2003.
 XX 28-MAR-2002; 2002EP-00007401.
 XX 05-NOV-2001; 2001JP-00379299.
 XX 25-JAN-2002; 2002US-00350978.
 XX (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX Tsugai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 WI; 2003-450961/43.
 DR N-PSDB; ADS62934.
 XX

XX Claim 1; Page; 222pp; English.
 XX The invention discloses a polynucleotide comprising a sequence selected
 CC from 1970 fully defined nucleotide sequences which encode novel
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 CC or its partial peptide, an antibody binding to the polypeptide or peptide
 CC of the polynucleotide, immunologically assaying the polypeptide or
 CC peptide of the polynucleotide by contacting the polypeptide or peptide
 CC with the antibody of the encoded protein, and observing the binding
 CC between the two, a transformant carrying the polynucleotide in an
 CC expressible manner and an antisense polynucleotide. The oligonucleotide
 CC is useful as a primer for synthesising the polynucleotide, or as a probe
 CC for detecting the polynucleotide. The polynucleotides and encoded
 CC proteins are useful as pharmaceutical agents and many disease-related
 CC genes may be included in them, for developing a diagnostic marker or
 CC medicines for regulation of their expression and activity, or as targets
 CC of gene therapy. The genes are involved in tissue and/or cell
 CC regeneration. Membrane proteins, signal transduction-related proteins,
 CC transcription-related proteins, disease-related proteins and genes
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
 CC neurological diseases, cancer, tumours). The CDNA may be used to regulate
 CC the activity or expression of the encoded protein to treat diseases. The
 CC sequence presented is a protein of the invention. Note: Some of the
 CC sequence data for this patent is not represented in the printed
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.

XX Sequence 463 AA;
 Query Match 16.4%; Score 628; DB 7; Length 463;
 Best Local Similarity 28.7%; Pred. No. 1.1e-31;
 Matches 162; Conservative 80; Mismatches 198; Indels 124; Gaps 16;

QY	151	LFAHGISHCPT--CDRRCQNGQCHDSESSIVYCVCPAGFTGSRCEHSQALHCHPEACG	208
Db	10	LSGADVGCSSGICDEASCIRHGCTAIKADSYICLCLPLG	50
QY	209	PDATCVNRDPGRGTYCRCHLGRSGLRCEGEVTTTPSL-SGAGSYLALPALTWHHEIR	265
Db	51	-----KGRHCEADATLITLTPQFRESLRSAATPWPPEPQHYLSF	88
QY	266	RLDVEFKLADPGLVLLFS---GGKSGVPEFVSLAMVGGHLEFRYELGSLAVLRSAEP	321

```

Db      89  MEFEITFRDSDGVLVLSYDTGSK-----DPLSINLAGHVEFRFCGSGTGVLRSEDP 143
QY      322  LALGRWHRVSAERLNKDGSLRVNGRVPYLRSSPGKSQGLNLHTLLYLGGVEPSVPLSPAT 381
Db      144  LTLGNWHELVRSTAKNGILOVKQKIVEGMAEGGFTQIKCNTDIFIGGVFNDDVKKNS 203
QY      382  NMSAHRGCVGVSVNGKRLDLYSFLGSGIGQCYDSSPCEROPCQHATCMPAGEVEF 441
Db      204  GVLKFFSGSIQKIIILNDRTHVKHDF--TSGVVENAAHPCVRAPCAHGSSCRPRKE--GY 260
QY      442  QCLCRDGFPGKDLCEHEENPCQLREFCLHGTCQGT--RCLCLPGFSGPRCQQSGHGIAE 499
Db      261  DCDCPLGPEG-----LH---COKAIIERAIPOFIG----- 288
QY      500  SDWHLGSGGNDAPGQYGAYFHDDGLAFPFGHVFSRSLPEVPETEIELEVRTSTASGLLLW 559
Db      289  -----RSLTYDNPDLKRVSG---SRS-----NVFWRFKTTAKDGLLLW 325
QY      560  QGVEVGEAGQKDFISLGLDGHLYFRYQLGSGEARLYSEDPINDGEWHRVTALREGRRG 619
Db      326  RG--DSPMRPNSDFISLGLRDLGALVFSYNLGSGVASIMVNGSFNDGRWHRVKAVRDQSG 383
QY      620  STQVDGEBELVSRSPGNVAVNAKGSVIYGAPDVATLTGGRFSSGITGCVKNLVLSAR 679
Db      384  KITVDYGARTGKSPGMRQLMINGALYVGGMKEIALHTNRQYMRGLVGCISHFTLST-- 441
QY      680  PGAPPPQLDQHRQAQANTRPC 703
Db      442  ----DYHISLVEDAVDGKNINTC 460

```

Search completed: March 9, 2004, 17:19:02
Job time : 91.8465 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:11:52 ; Search time 6.86906 Seconds
(without alignments)
2152.832 Million cell updates/sec

Title: US-10-006-011A-9

Perfect score: 1566

Sequence: 1 CERQPCQHGATCMEAGEYEF.....QLDLQHRQAQAGANTRPCPS 284

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1566	100.0	4391	1	PGBM_HUMAN
2	1370.5	87.5	3707	1	PGBM_MOUSE
3	349	22.3	3375	1	UN52_CABEL
4	333.5	21.3	1959	1	AGRI_RAT
5	314.5	20.1	1328	1	AGRI_DISOM
6	301.5	19.3	1955	1	AGRI_CHICK
7	269	17.2	5147	1	FAT_DROME
8	253.5	16.2	1529	1	SLT2_HUMAN
9	242	15.5	4705	1	FAT2_DROME
10	225	14.4	3579	1	SPAN_DROME
11	223.5	14.3	2110	1	LMA2_HUMAN
12	214	13.7	1578	1	NK3A_RAT
13	211.5	13.5	2923	1	CLR2_HUMAN
14	207.5	13.3	2144	1	CLR2_RAT
15	207.5	13.3	3106	1	LMA2_MOUSE
16	206.5	13.2	3312	1	CLR3_HUMAN
17	204.5	13.1	1712	1	NK2A_HUMAN
18	204.5	13.1	3313	1	CLR3_RAT
19	200	12.8	1541	1	NK3A_HUMAN
20	199	12.7	2321	1	NTC3_HUMAN
21	199	12.7	3014	1	CLR1_HUMAN
22	197.5	12.6	1530	1	NK1A_BOVIN
23	194.5	12.4	1715	1	NK2A_RAT
24	193	12.3	1363	1	NK1A_CHICK
25	193	12.3	1477	1	NK1A_HUMAN
26	193	12.3	1514	1	NK1A_RAT
27	191	12.2	3034	1	CLR1_MOUSE
28	190.5	12.2	618	1	DLL3_HUMAN
29	189.5	12.1	3301	1	CLR3_MOUSE
30	188.5	12.0	3075	1	LMA1_HUMAN
31	188	11.8	2215	1	CDN2_DROME
32	185.5	11.8	3672	1	LML2_CABEL
33	184	11.7	3097	1	CADN_DROME

34	183	11.7	1504	1	SLIT_DROME
35	182.5	11.7	2003	1	NTC4_HUMAN
36	182	11.6	2318	1	NTC3_MOUSE
37	181.5	11.6	2531	1	NTC1_MOUSE
38	181.5	11.6	2531	1	NTC1_RAT
39	180.5	11.5	589	1	DLL3_RAT
40	180.5	11.5	592	1	DLL3_MOUSE
41	180.5	11.5	1247	1	JAG2_MOUSE
42	180.5	11.5	1321	1	PGCN_HUMAN
43	180	11.5	1238	1	JAG2_HUMAN
44	179.5	11.5	1202	1	JAG2_RAT
45	179.5	11.5	1242	1	JAG1_BRARE

ALIGNMENTS

RESULT 1

PGBM_HUMAN

ID PGBM_HUMAN STANDARD; PRT; 4391 AA.

AC P981F0; Q16287; Q9H3V5;

DT 01-OCT-1996 (Rel. 34, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Basement membrane-specific heparan sulfate proteoglycan core

DE protein precursor (HSPG) (Perlecan) (PLC).

OS HSPG2.

OC Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RX MEDLINE=92112994; PubMed=1730768;

RA Kallunki P., Tryggvason K.

RT "Human basement membrane heparan sulfate proteoglycan core protein: a

467-kD protein containing multiple domains resembling elements of the

low density lipoprotein receptor, laminin, neural cell adhesion

molecules, and epidermal growth factor.";

RL J. Cell Biol. 116:559-571(1992).

RP SEQUENCE FROM N.A.

RT TISSUE=Colon, and Skin;

RC MEDLINE=92235084; PubMed=1569102;

RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;

"Primary structure of the human heparan sulfate proteoglycan from

basement membrane (HSPG2/perlecan). A chimeric molecule with multiple

domains homologous to the low density lipoprotein receptor, laminin,

neural cell adhesion molecules, and epidermal growth factor.";

RL J. Biol. Chem. 267:8544-8557(1992).

RP SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TYR-1532.

RX MEDLINE=20553141; PubMed=1101850;

RA Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,

Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,

Ramson D., Urtizbera J.A., Lehmann-Horn F., Weissenbach J.,

Kentati F., Fontaine B.;

"Perlecan, the major proteoglycan of basement membranes, is altered in

patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).";

Nat. Genet. 26:480-483(2000).

RP SEQUENCE OF 1016-1470 FROM N.A.

RT TISSUE=Colon;

RX MEDLINE=91365376; PubMed=1679749;

RA Dodge G.R., Kovalesky I., Chu M.L., Hassell J.R., McBride O.W.,

Xi H.F., Iozzo R.V.;

"Heparan sulfate proteoglycan of human colon: partial molecular

cloning, cellular expression, and mapping of the gene (HSPG2) to the

short arm of human chromosome 1.";

RL Genomics 10:673-680(1991).

RP SEQUENCE OF 890-1396 FROM N.A.

RX TISSUE=Fibrosarcoma;

RX MEDLINE=92120660; PubMed=1685141;
 RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
 RA Tryggvason K.;
 RT "Cloning of human heparan sulfate proteoglycan core protein,
 RT assignment of the gene (HSPG2) to lp36.1--p35 and identification of
 RT a BamHI restriction fragment length polymorphism.";
 RL Genomics 11:389-396(1991).
 RN [6]
 RP SEQUENCE OF 1-21 FROM N.A.
 RX MEDLINE=94052171; PubMed=8234307;
 RA Cohen I.R., Graessel S., Murodoch A.D., Iozzo R.V.;
 RA "Structural characterization of the complete human perlecan gene and
 RT its promoter.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
 RN [7]
 RP CARBOHYDRATE-LINKAGE SITE ASN-2121.
 RX MEDLINE=22660472; PubMed=12754519;
 RA Zhang H., Li X.-J., Martin D.B., Aebbersold R.;
 RA "Identification and quantification of N-linked glycoproteins using
 RT hydrazide chemistry, stable isotope labeling and mass spectrometry.";
 RL Nat. Biotechnol. 21:660-666(2003).
 CC !- FUNCTION: This protein is an integral component of basement
 CC membranes. It is responsible for the fixed negative electrostatic
 CC charge and is involved in the charge-selective ultrafiltration
 CC properties. It serves as an attachment substrate for cells.
 CC !- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
 CC dimers or stellate structures. It interacts with other basement
 CC membrane components such as laminin, prolargin and collagen type
 CC IV.
 CC !- SUBCELLULAR LOCATION: Extracellular.
 CC !- TISSUE SPECIFICITY: Found in the basement membranes.
 CC !- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
 CC AND O-LINKED OLIGOSACCHARIDES.
 CC !- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
 CC syndrome (SJS1) [MIM:255800]; a rare autosomal recessive disorder
 CC characterized by permanent myotonia (prolonged failure of muscle
 CC relaxation) and skeletal dysplasia, resulting in reduced stature,
 CC kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
 CC !- SIMILARITY: Contains 4 LDL-receptor class A domains.
 CC !- SIMILARITY: Contains 11 laminin EGF-like domains.
 CC !- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
 CC !- SIMILARITY: Contains 3 laminin G-like domains.
 CC !- SIMILARITY: Contains 4 EGF-like domains.
 CC !- SIMILARITY: Contains 1 SEA domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X62515; CAA44373.1; --
 DR EMBL; M85289; AA82700.1; --
 DR EMBL; AL445795; CAC18534.1; --
 DR EMBL; M64283; AA852639.1; --
 DR EMBL; S76436; AAB21121.2; --
 DR EMBL; L22078; -- NOT_ANNOTATED_CDS.
 DR PIR; A38096; A38096.
 DR HSPG; P00740; LEDM.
 DR Siena-2DPAGE; P98160; --
 DR Genew; HGNC:5273; HSPG2.
 DR MIM; 142461; --
 DR MIM; 255800; --
 DR InterPro; IPR008985; ConA_like Lec_gl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003598; Ig_c2.

DR InterPro; IPR003596; Ig_v.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF00008; EGF_4.
 DR Pfam; PF00047; ig_22.
 DR Pfam; PF00052; laminin_B_3.
 DR Pfam; PF00053; laminin_EGF_7.
 DR Pfam; PF00054; laminin_G_3.
 DR Pfam; PF00057; ldl_recept_a_4.
 DR Pfam; PF01390; SEA_1.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR ProDom; PD003031; Laminin_B_3.
 DR SMART; SM01181; EGF_15_12.
 DR SMART; SM01180; EGF_Lam_12.
 DR SMART; SM00409; IG_22.
 DR SMART; SM00408; ICG2_21.
 DR SMART; SM00406; IGV_7.
 DR SMART; SM00281; LamB_3.
 DR SMART; SM00282; LamG_3.
 DR SMART; SM00192; LDLa_4.
 DR SMART; SM00200; SEA_1.
 DR PROSITE; PS00022; EGF_1_9.
 DR PROSITE; PS01186; EGF_2_6.
 DR PROSITE; PS00026; EGF_3_4.
 DR PROSITE; PS50835; IG_LIKE_22.
 DR PROSITE; PS50025; Lam_G_DOMAIN_3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF_11.
 DR PROSITE; PS01209; LDLRA_1_4.
 DR PROSITE; PS50068; LDLRA_2_4.
 DR PROSITE; PS50024; SEA_1.
 DR PROSITE; PS50024; SEA_1.
 KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
 KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
 KW Extracellular matrix; EGF-like domain; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 22 4391
 FT DOMAIN 80 194
 FT DOMAIN 198 235
 FT DOMAIN 284 320
 FT DOMAIN 324 360
 FT DOMAIN 367 404
 FT DOMAIN 405 504
 FT DOMAIN 521 530
 FT DOMAIN 531 730
 FT DOMAIN 731 763
 FT DOMAIN 764 813
 FT DOMAIN 814 871
 FT DOMAIN 879 923
 FT DOMAIN 924 933
 FT DOMAIN 934 1123
 FT DOMAIN 1126 1158
 FT DOMAIN 1159 1208
 FT DOMAIN 1209 1265
 FT DOMAIN 1275 1324
 FT DOMAIN 1325 1334
 FT DOMAIN 1335 1529
 FT DOMAIN 1530 1562
 FT DOMAIN 1563 1612
 FT DOMAIN 1613 1670
 FT DOMAIN 1671 1771
 FT DOMAIN 1772 1865
 FT DOMAIN 1866 1955
 FT DOMAIN 1956 2051
 FT DOMAIN 2052 2151
 FT DOMAIN 2152 2244
 FT DOMAIN 2245 2340
 FT DOMAIN 2341 2436
 FT DOMAIN 2437 2533
 FT DOMAIN 2534 2629
 FT DOMAIN 2630 2726
 LDL-RECEPTOR CLASS A 1.
 LDL-RECEPTOR CLASS A 2.
 LDL-RECEPTOR CLASS A 3.
 LDL-RECEPTOR CLASS A 4.
 IG-LIKE C2-TYPE 1
 LAMININ EGF-LIKE 1 (N-TERMINAL).
 LAMININ EGF-LIKE 1 (C-TERMINAL).
 LAMININ EGF-LIKE 2.
 LAMININ EGF-LIKE 3.
 LAMININ EGF-LIKE 4 (INCOMPLETE).
 LAMININ EGF-LIKE 5 (N-TERMINAL).
 LAMININ EGF-LIKE 5 (C-TERMINAL).
 LAMININ EGF-LIKE 6.
 LAMININ EGF-LIKE 7.
 LAMININ EGF-LIKE 8.
 LAMININ EGF-LIKE 9 (N-TERMINAL).
 LAMININ EGF-LIKE 9 (C-TERMINAL).
 LAMININ EGF-LIKE 10.
 LAMININ EGF-LIKE 11.
 IG-LIKE C2-TYPE 2.
 IG-LIKE C2-TYPE 3.
 IG-LIKE C2-TYPE 4.
 IG-LIKE C2-TYPE 5.
 IG-LIKE C2-TYPE 6.
 IG-LIKE C2-TYPE 7.
 IG-LIKE C2-TYPE 8.
 IG-LIKE C2-TYPE 9.
 IG-LIKE C2-TYPE 10.
 IG-LIKE C2-TYPE 11.
 IG-LIKE C2-TYPE 12.

Query Match 100.0%; Score 1566; DB 1; Length 4391;
 Best Local Similarity 100.0%; Pred. No. 2e-103;
 Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CERPCQHGATCMFAGYEFQCLCRDGFKGLDCEHENPCQLREPCCLHGCTCGQTRCLCL 60
 DB 4108 CERPCQHGATCMFAGYEFQCLCRDGFKGLDCEHENPCQLREPCCLHGCTCGQTRCLCL 4167

QY 61 PFGSPRCQOQSGHGIAEBSWHLESGSGNDAPQYGFYFDDGFLAPPHVFRSLPEVP 120
 DB 4168 PFGSPRCQOQSGHGIAEBSWHLESGSGNDAPQYGFYFDDGFLAPPHVFRSLPEVP 4227

QY 121 ETILEVRTSTASGILLWQGVGEAGQKDFISLGLQDGHVFRYQLSGEARLYSEDP 180
 DB 4228 ETILEVRTSTASGILLWQGVGEAGQKDFISLGLQDGHVFRYQLSGEARLYSEDP 4287

QY 181 INDGEHRVTRALREGRGSIQVDGEELVSGRSPGNVAVNAKGSVYTGAPDVAITLGG 240
 DB 4288 INDGEHRVTRALREGRGSIQVDGEELVSGRSPGNVAVNAKGSVYTGAPDVAITLGG 4347

QY 241 FSSGILTCVKNLVLSARPGAPPPQPLDQHRAGAGANTRPCPS 284
 DB 4348 FSSGILTCVKNLVLSARPGAPPPQPLDQHRAGAGANTRPCPS 4391

RESULT 2

PCBM_MOUSE STANDARD; PRT; 3707 AA.

AC 005793;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Basement membrane-specific heparan sulfate proteoglycan core protein precursor (HSPG) (perlecan) (PLC).
 GN HSPG2.
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Melanoma;
 EX MEDLINE=92078153; PubMed=1744087;
 RA Noonan D.M., Fulle A., Valente P., Cai S., Horigan E., Sasaki M., Yamada Y., Hassell J.R.;
 RA "The complete sequence of perlecan, a basement membrane heparan sulfate proteoglycan, reveals extensive similarity with laminin A chain, low density lipoprotein-receptor, and the neural cell adhesion molecule";
 RL J. Biol. Chem. 266:22939-22947(1991).
 RN [2]
 RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=89034110; PubMed=2372708;
 RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M., Yamada Y., Hassell J.R.;
 RT "Identification of cDNA clones encoding different domains of the basement membrane heparan sulfate proteoglycan";
 RL J. Biol. Chem. 263:16379-16387(1988).
 CC -1- FUNCTION: This protein is an integral component of basement membranes. It is responsible for the fixed negative electrostatic charge and is involved in the charge-selective ultrafiltration properties. It serves as an attachment substrate for cells.
 CC -1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in dimers or stellate structures. It interacts with other basement membrane components such as laminin, prolargin and collagen type IV.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: Found in the basement membranes.
 CC -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.
 CC -1- SIMILARITY: Contains 11 laminin EGF-like domains.

CC -1- SIMILARITY: Contains 3 laminin IV domains.
 CC -1- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 3 laminin G-like domains.
 CC -1- SIMILARITY: Contains 1 EGF-like domain.
 CC -1- SIMILARITY: Contains 1 SEA domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M71174; AAA39911.1;
 DR EMBL; J04054; AAA39899.1;
 DR EMBL; J04055; AAA39912.1;
 DR FIR; S18252; S18252.
 DR PDB; 1GL4; 28-NOV-01.
 DR MGD; MGI:96257; Hspg2.
 DR GO; GO:0005604; C:basement membrane; IDA.
 DR GO; GO:0008104; P:protein localization; IMP.
 DR InterPro; IPR008985; Consa_like_lec_g1.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_C2.
 DR InterPro; IPR000034; Laminin B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00047; Ig; 15.
 DR Pfam; PF00052; laminin_B; 3.
 DR Pfam; PF00053; laminin_EGF; 7.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF00057; ldl_recept_a; 4.
 DR Pfam; PF01390; SEA; 1.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR ProDom; PD003031; Laminin B; 3.
 DR SMART; SM00180; EGF_Lam; 7.
 DR SMART; SM00408; IGC2; 14.
 DR SMART; SM00281; LamB; 3.
 DR SMART; SM00282; LamG; 3.
 DR SMART; SM00192; LDLa; 4.
 DR SMART; SM00200; SEA; 1.
 DR PROSITE; PS00022; EGF_1; 8.
 DR PROSITE; PS01186; EGF_2; 5.
 DR PROSITE; PS50026; EGF_3; 4.
 DR PROSITE; PS50835; IG_LIKE; 15.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 DR PROSITE; PS01209; LDLRA_1; 4.
 DR PROSITE; PS50068; LDLRA_2; 4.
 DR PROSITE; PS50024; SEA; 1.
 KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
 KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
 KW Extracellular matrix; EGF-like domain; 3D-structure.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 3707 BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE PROTEIN.
 FT FT DOMAIN 80 194 SEA.
 FT FT DOMAIN 195 234 LDL-RECEPTOR CLASS A 1.
 FT FT DOMAIN 281 319 LDL-RECEPTOR CLASS A 2.
 FT FT DOMAIN 320 359 LDL-RECEPTOR CLASS A 3.
 FT FT DOMAIN 360 404 LDL-RECEPTOR CLASS A 4.
 FT FT DOMAIN 404 503 IG-LIKE C2-TYPE 1.
 FT FT DOMAIN 521 530 LAMININ EGF-LIKE 1 (N-TERMINAL).
 FT FT DOMAIN 531 730 LAMININ DOMAIN IV 1 (DOMAIN III A).
 FT FT DOMAIN 731 763 LAMININ EGF-LIKE 1 (C-TERMINAL).
 FT FT DOMAIN 764 813 LAMININ EGF-LIKE 2.
 FT FT DOMAIN 814 871 LAMININ EGF-LIKE 3.

FT DOMAIN 879 923 LAMININ EGF-LIKE 4 (INCOMPLETE).
 FT DOMAIN 924 933 LAMININ EGF-LIKE 5 (N-TERMINAL).
 FT DOMAIN 934 1125 LAMININ DOMAIN IV 2 (DOMAIN III B).
 FT DOMAIN 1126 1158 LAMININ EGF-LIKE 5 (C-TERMINAL).
 FT DOMAIN 1159 1208 LAMININ EGF-LIKE 6.
 FT DOMAIN 1209 1265 LAMININ EGF-LIKE 7.
 FT DOMAIN 1275 1324 LAMININ EGF-LIKE 8.
 FT DOMAIN 1325 1334 LAMININ EGF-LIKE 9 (N-TERMINAL).
 FT DOMAIN 1335 1529 LAMININ DOMAIN IV 3 (DOMAIN III C).
 FT DOMAIN 1530 1562 LAMININ EGF-LIKE 9 (C-TERMINAL).
 FT DOMAIN 1563 1612 LAMININ EGF-LIKE 10.
 FT DOMAIN 1613 1670 LAMININ EGF-LIKE 11.
 FT DOMAIN 1671 1771 IG-LIKE C2-TYPE 2.
 FT DOMAIN 1772 1865 IG-LIKE C2-TYPE 3.
 FT DOMAIN 1866 1954 IG-LIKE C2-TYPE 4.
 FT DOMAIN 1955 2049 IG-LIKE C2-TYPE 5.
 FT DOMAIN 2050 2148 IG-LIKE C2-TYPE 6.
 FT DOMAIN 2149 2244 IG-LIKE C2-TYPE 7.
 FT DOMAIN 2245 2343 IG-LIKE C2-TYPE 8.
 FT DOMAIN 2344 2436 IG-LIKE C2-TYPE 9.
 FT DOMAIN 2437 2532 IG-LIKE C2-TYPE 10.
 FT DOMAIN 2533 2619 IG-LIKE C2-TYPE 11.
 FT DOMAIN 2620 2720 IG-LIKE C2-TYPE 12.
 FT DOMAIN 2721 2809 IG-LIKE C2-TYPE 13.
 FT DOMAIN 2810 2895 IG-LIKE C2-TYPE 14.
 FT DOMAIN 2896 2980 IG-LIKE C2-TYPE 15.
 FT DOMAIN 2984 3162 LAMININ G-LIKE 1.
 FT DOMAIN 3163 3241 EGF-LIKE.
 FT DOMAIN 3245 3425 LAMININ G-LIKE 2.
 FT DOMAIN 3518 3705 LAMININ G-LIKE 3.
 FT SITE 65 67 HEPARAN SULFATE (POTENTIAL).
 FT SITE 71 73 HEPARAN SULFATE (POTENTIAL).
 FT SITE 76 78 HEPARAN SULFATE (POTENTIAL).
 FT SITE 3615 3617 MEDIATES MOTOR NEURON ATTACHMENT (POTENTIAL).
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 206 225 BY SIMILARITY.
 FT DISULFID 229 234 BY SIMILARITY.
 FT DISULFID 285 297 BY SIMILARITY.
 FT DISULFID 292 310 BY SIMILARITY.
 FT DISULFID 304 319 BY SIMILARITY.
 FT DISULFID 325 337 BY SIMILARITY.
 FT DISULFID 332 350 BY SIMILARITY.
 FT DISULFID 344 359 BY SIMILARITY.
 FT DISULFID 368 381 BY SIMILARITY.
 FT DISULFID 375 394 BY SIMILARITY.
 FT DISULFID 388 403 BY SIMILARITY.
 FT DISULFID 428 479 BY SIMILARITY.
 FT DISULFID 764 773 BY SIMILARITY.
 FT DISULFID 766 780 BY SIMILARITY.
 FT DISULFID 783 792 BY SIMILARITY.
 FT DISULFID 795 811 BY SIMILARITY.
 FT DISULFID 814 829 BY SIMILARITY.
 FT DISULFID 816 839 BY SIMILARITY.
 FT DISULFID 842 851 BY SIMILARITY.
 FT DISULFID 854 869 BY SIMILARITY.
 FT DISULFID 1159 1168 BY SIMILARITY.
 FT DISULFID 1161 1175 BY SIMILARITY.
 FT DISULFID 1178 1187 BY SIMILARITY.
 FT DISULFID 1190 1206 BY SIMILARITY.
 FT DISULFID 1209 1224 BY SIMILARITY.
 FT DISULFID 1211 1234 BY SIMILARITY.
 FT DISULFID 1237 1246 BY SIMILARITY.
 FT DISULFID 1249 1263 BY SIMILARITY.
 FT DISULFID 1275 1287 BY SIMILARITY.
 FT DISULFID 1277 1293 BY SIMILARITY.
 FT DISULFID 1295 1304 BY SIMILARITY.
 FT DISULFID 1307 1322 BY SIMILARITY.
 FT DISULFID 1563 1572 BY SIMILARITY.
 FT DISULFID 1565 1579 BY SIMILARITY.
 FT DISULFID 1582 1591 BY SIMILARITY.
 FT DISULFID 1594 1610 BY SIMILARITY.
 FT DISULFID 1613 1628 BY SIMILARITY.

FT DISULFID 1615 1638 BY SIMILARITY.
 FT DISULFID 1641 1650 BY SIMILARITY.
 FT DISULFID 1653 1668 BY SIMILARITY.
 FT DISULFID 1792 1839 BY SIMILARITY.
 FT DISULFID 1886 1932 BY SIMILARITY.
 FT DISULFID 2076 2021 BY SIMILARITY.
 FT DISULFID 2073 2118 BY SIMILARITY.
 FT DISULFID 2170 2215 BY SIMILARITY.
 FT DISULFID 2268 2313 BY SIMILARITY.
 Query Match 87.5%; Score 1370.5; DB 1; Length 3707;
 Best Local Similarity 86.3%; Pred. No. 1.4e-89;
 Matches 245; Conservative 20; Mismatches 18; Indels 1; Gaps 1;
 QY 1 CERPCQHGATCMPAGEYEFQCLCRDQYKGLCEHEENPCQLREPCCLHGCTCGTRCLCL 60
 Db 3425 CERPCRNATCMPAGEYEFQCLCRDQYKGLCEHEENPCQLREPCCLHGCTCGTRCLCL 3484
 QY 61 PGFSGRCQQGGHGAESDWHLEGGGNDARFGQYGFDDGFLAFPGHVFSSLSPEVP 120
 Db 3485 PGFSGRCQQGGHGAESDWHLEGGGNDARFGQYGFDDGFLAFPGHVFSSLSPEVP 3544
 QY 121 ETIELEVRSTASGLLLWQGVVEGAGQKDISLGLDGHVFRYQLGSGEARLVSEDP 180
 Db 3545 ETIEFEVRYSTADGLLLWQGV-VREASRSKDISLGLDGHVFRYQLGSGEARLVSEDP 3603
 QY 181 INDGEWHRYTALREGRGSIQVDGELVSGRPGPNVAVNAKGSVYVIGGAPDVATITGGR 240
 Db 3604 INDGEWHRYTALREGRGSIQVDGELVSGRPGPNVAVNAKGSVYVIGGAPDVATITGGR 3663
 QY 241 FSSGITCVNVLVHSARGAPPQPLDLQHRQAQAGANTRPCPS 284
 Db 3664 FSSGITCVNVLVHSARGAPPQPLDLQHRQAQAGANTRPCPS 3707
 RESULT 3
 UN52_CABEEL STANDARD; PRT; 3375 AA.
 ID_UN52_CABEEL Q06561; O18263; Q9XTD2; Q9XTI5;
 AC Q06561; O18263; Q9XTD2; Q9XTI5;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Basement membrane proteoglycan precursor (Perlecan homolog)
 DE (Uncoordinated protein 52).
 GN UNC-52 OR ZC101.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 CX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM A), AND FUNCTION.
 RX MEDLINE=9339574; PubMed=8393416;
 RA Rogalski T.M., Williams B.D., Mullen G.P., Moerman D.G.;
 RT "Products of the unc-52 gene in Caenorhabditis elegans are homologous
 RT to the core protein of the mammalian basement membrane heparan
 RT sulfate proteoglycan."
 RL Genes Dev. 7:1471-1484(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2.
 RA Percy C.M., Baynes C.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RA Durbin R.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Probable role in myofibril assembly and/or attachment
 CC of the myofibril lattice to the cell membrane. May be an
 CC extracellular anchor for integrin receptors in muscle.
 CC -!- SUBCELLULAR LOCATION: Extracellular matrix.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=4;
 CC Name=;

CC IsoId=Q06561-1; Sequence=Displayed;
 CC Note=No experimental confirmation available;
 CC Name=a;
 CC IsoId=Q06561-2; Sequence=VSP_007195, VSP_007196;
 CC Name=b;
 CC IsoId=Q06561-3; Sequence=VSP_007191, VSP_007192;
 CC Note=No experimental confirmation available;
 CC Name=c;
 CC IsoId=Q06561-4; Sequence=VSP_007193, VSP_007194, VSP_007195,
 CC VSP_007196;

CC Note=No experimental confirmation available;
 CC TISSUE SPECIFICITY: Found in the basement membrane of all
 CC contractile tissues. It is concentrated over muscle dense bodies
 CC and M-lines which are associated with beta-integrin.
 CC -!- DEVELOPMENTAL STAGE: Synthesized early in embryogenesis.
 CC -!- SIMILARITY: Contains 3 LDL-receptor class A domains.
 CC -!- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.
 CC -!- SIMILARITY: Contains 7 laminin EGF-like domains.
 CC -!- SIMILARITY: Contains 3 laminin G-like domains.
 CC -!- SIMILARITY: Contains 2 laminin IV domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to licenses@isb-sib.ch).
 CC -----

CC EMBL; L13458; AAA28156.1; -
 DR EMBL; Z93375; CAB07567.1; -
 DR EMBL; Z93395; CAB07567.1; JOINED.
 DR EMBL; Z93375; CAB07568.1; -
 DR EMBL; Z93395; CAB07568.1; JOINED.
 DR EMBL; Z93375; CAB07569.1; -
 DR EMBL; Z93395; CAB07569.1; JOINED.
 DR EMBL; Z93395; CAB07704.1; -
 DR EMBL; Z93395; CAB07704.1; -
 DR EMBL; Z93395; CAB07706.1; -
 DR EMBL; Z93375; CAB07706.1; JOINED.
 DR EMBL; Z93395; CAB07707.1; -
 DR EMBL; Z93375; CAB07707.1; JOINED.
 DR EMBL; Z93395; CAB07708.1; -
 DR EMBL; Z93375; CAB07708.1; JOINED.
 DR HSP; P01130; ILDR.
 DR WormPep; ZC101.2a; CE15028.
 DR WormPep; ZC101.2b; CE15030.
 DR WormPep; ZC101.2c; CE15034.
 DR WormPep; ZC101.2e; CE18424.
 DR GO; GO:0005578; C:extracellular matrix; IEP.
 DR GO; GO:0030239; P:myofibril assembly; IEP.
 DR InterPro; IPR008985; Cona_like_rec_g1.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR Pfam; PF00047; ig; 16
 DR Pfam; PF00052; laminin_B; 2.
 DR Pfam; PF00053; laminin_EGF; 5.
 DR Pfam; PF00057; ldl_recept_a; 3.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR ProDom; PD003031; Laminin_B; 2.
 DR SMART; SMC0181; EGF; 6.
 DR SMART; SMC0179; EGF_CA; 2.
 DR SMART; SMC0180; EGF_Lam; 6.
 DR SMART; SMC0409; IG; 17.
 DR SMART; SMC0408; IGC2; 17.
 DR SMART; SMC0281; LamB; 1.
 DR SMART; SMC0281; LamB; 1.

DR SMART; SMC0282; LamG; 3.
 DR SMART; SMC0192; LDLa; 3.
 DR PROSITE; PS00022; EGF_1; 7.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS00026; EGF_3; 3.
 DR PROSITE; PS00835; IG_LIKE; 17.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
 DR PROSITE; PS01209; LDLa_1; 3.
 DR PROSITE; PS00681; LDLa_2; 3.
 KW Proteoglycan; Glycoprotein; Signal; Immunoglobulin domain; Repeat;
 KW Basement membrane; Extracellular matrix; Alternative splicing;
 KW Laminin EGF-like domain.
 FT SIGNAL; 23 3375 POTENTIAL.
 FT CHAIN; 23 3375 BASEMENT MEMBRANE PROTEOGLYCAN.
 FT DOMAIN; 45 130 IG-LIKE C2-TYPE 1.
 FT DOMAIN; 148 184 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN; 189 225 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN; 232 269 LDL-RECEPTOR CLASS A 3.
 FT DOMAIN; 271 355 IG-LIKE C2-TYPE 2.
 FT DOMAIN; 384 431 LAMININ EGF-LIKE 1 (INCOMPLETE).
 FT DOMAIN; 442 441 LAMININ EGF-LIKE 2 (N-TERMINAL).
 FT DOMAIN; 442 633 LAMININ DOMAIN IV 1.
 FT DOMAIN; 634 666 LAMININ EGF-LIKE 2 (C-TERMINAL).
 FT DOMAIN; 674 720 LAMININ EGF-LIKE 3 (INCOMPLETE).
 FT DOMAIN; 721 730 LAMININ EGF-LIKE 4 (N-TERMINAL).
 FT DOMAIN; 731 921 LAMININ DOMAIN IV 2.
 FT DOMAIN; 922 954 LAMININ EGF-LIKE 4 (C-TERMINAL).
 FT DOMAIN; 955 1004 LAMININ EGF-LIKE 5.
 FT DOMAIN; 1011 1060 LAMININ EGF-LIKE 6.
 FT DOMAIN; 1061 1111 LAMININ EGF-LIKE 7.
 FT DOMAIN; 1126 1222 IG-LIKE C2-TYPE 3.
 FT DOMAIN; 1226 1311 IG-LIKE C2-TYPE 4.
 FT DOMAIN; 1319 1401 IG-LIKE C2-TYPE 5.
 FT DOMAIN; 1410 1499 IG-LIKE C2-TYPE 6.
 FT DOMAIN; 1503 1585 IG-LIKE C2-TYPE 7.
 FT DOMAIN; 1588 1680 IG-LIKE C2-TYPE 8.
 FT DOMAIN; 1690 1785 IG-LIKE C2-TYPE 9.
 FT DOMAIN; 1793 1878 IG-LIKE C2-TYPE 10.
 FT DOMAIN; 1886 1970 IG-LIKE C2-TYPE 11.
 FT DOMAIN; 1973 2069 IG-LIKE C2-TYPE 12.
 FT DOMAIN; 2073 2263 IG-LIKE C2-TYPE 13.
 FT DOMAIN; 2173 2260 IG-LIKE C2-TYPE 14.
 FT DOMAIN; 2263 2343 IG-LIKE C2-TYPE 15.
 FT DOMAIN; 2349 2435 IG-LIKE C2-TYPE 16.
 FT DOMAIN; 2446 2530 IG-LIKE C2-TYPE 17.
 FT DOMAIN; 2532 2713 LAMININ G-LIKE 1.
 FT DOMAIN; 2793 2960 LAMININ G-LIKE 2.
 FT DOMAIN; 2961 3093 GLU-RICH.
 FT DOMAIN; 2972 3066 THR-RICH.
 FT DOMAIN; 3180 3359 LAMININ G-LIKE 3.
 FT DISULFID; 66 114 BY SIMILARITY.
 FT DISULFID; 149 161 BY SIMILARITY.
 FT DISULFID; 156 174 BY SIMILARITY.
 FT DISULFID; 168 183 BY SIMILARITY.
 FT DISULFID; 190 202 BY SIMILARITY.
 FT DISULFID; 197 215 BY SIMILARITY.
 FT DISULFID; 209 224 BY SIMILARITY.
 FT DISULFID; 233 246 BY SIMILARITY.
 FT DISULFID; 240 259 BY SIMILARITY.
 FT DISULFID; 253 268 BY SIMILARITY.
 FT DISULFID; 955 964 BY SIMILARITY.
 FT DISULFID; 957 971 BY SIMILARITY.
 FT DISULFID; 974 983 BY SIMILARITY.
 FT DISULFID; 986 1002 BY SIMILARITY.
 FT DISULFID; 1011 1021 BY SIMILARITY.
 FT DISULFID; 1013 1027 BY SIMILARITY.
 FT DISULFID; 1030 1039 BY SIMILARITY.
 FT DISULFID; 1042 1058 BY SIMILARITY.
 FT DISULFID; 1061 1069 BY SIMILARITY.
 FT DISULFID; 1063 1079 BY SIMILARITY.
 FT DISULFID; 1082 1091 BY SIMILARITY.
 FT DISULFID; 1094 1109 BY SIMILARITY.

```

FT DISULFID 1152 1200 BY SIMILARITY.
FT DISULFID 1384 1384 BY SIMILARITY.
FT DISULFID 1435 1481 BY SIMILARITY.
FT DISULFID 1527 1573 BY SIMILARITY.
FT DISULFID 1618 1663 BY SIMILARITY.
FT DISULFID 1719 1767 BY SIMILARITY.
FT DISULFID 1814 1861 BY SIMILARITY.
FT DISULFID 1907 1954 BY SIMILARITY.
FT DISULFID 1988 2053 BY SIMILARITY.
FT DISULFID 2099 2147 BY SIMILARITY.
FT DISULFID 2195 2242 BY SIMILARITY.
FT DISULFID 2284 2329 BY SIMILARITY.
FT DISULFID 2374 2420 BY SIMILARITY.
FT DISULFID 2467 2514 BY SIMILARITY.
FT CARBOHYD 1422 1422 N-LINKED (GLCNAC . . ) (POTENTIAL).
FT CARBOHYD 2476 2476 N-LINKED (GLCNAC . . ) (POTENTIAL).
FT CARBOHYD 2950 2950 N-LINKED (GLCNAC . . ) (POTENTIAL).
FT CARBOHYD 3143 3143

Query Match          22.3%; Score 349; DB 1; Length 3375;
Best Local Similarity 31.2%; Pred. No. 6.5e-17;
Matches 81; Conservative 41; Mismatches 102; Indels 36; Gaps 6;

QY 30 GDLCHEH--NPGQRLEPLHGTCQGR-----CLCLPFGSPRCQQGGSHGIAESDWH 82
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 3129 GDVYSTQEPNNIC-ANSTCGMNGCVPRNTHYTCECKLYYDGTCSLFK----- 3177
QY 83 LEGSGNDPAGGYGFHDDGFAPPGHVFGRSLPEVPTLELVRTSTAGLILWQVE 142
. 3178 -----PEHAARFDGDFAFIELSDDEFPHLISEKDEIVAFKFKTEQQVLLWQSG-Q 3227
QY 143 VGEAGQGKDFSLGQDGHVFRYVLGSGEARLVSEDPNDGQVHRVYALREGRRGSIQV 202
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 3228 RPTVQMEDYISVGVNGLHFSLFVSGVAVAAHLSERVDGKHSVRFERKREGQMRI 3287
QY 203 DGEELVSRSPGPVAVNAKSVVIGGAPVDVATLTGGRFFSGITGCVKNIVLHSARPGAP 262
. 3288 DNYREVDGRGTGLMAMVNDGNFVGGVPDISKATGGLFNNFVGGIADVELNGVK----- 3343
QY 263 PPQPLDLQHRQAQAGANTRPC 282
. 3344 ----LDLMTAIDGNVRFC 3359
          STANDARD;          PRT; 1959 AA.

RESULT 4
AGRI_RAT
AC Q63034; Q63034;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Agrin precursor.
GN AGRN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
RC TISSUE=embryonic spinal cord;
RX MEDLINE=9122570; PubMed=1851019;
RA Rupp F., Payan D.G., Magill-Solc C., Cowan D.M., Scheller R.H.;
RT "Structure and chromosomal localization of the mammalian agrin gene.";
RL Neuron 6:811-823(1991).
RN [2]
RP SEQUENCE OF 1777-1801 FROM N.A.
RX MEDLINE=92407628; PubMed=1326508;
RA Rupp F., Cezeclik T., Linial M., Peterson K., Francke U., Scheller R.;
RT "Structure and chromosomal localization of the mammalian agrin gene.";
RL J. Neurosci. 12:3535-3544(1992).
CC -!- FUNCTION: Component of the basal lamina that causes the
CC aggregation of acetylcholine receptors and acetylcholine-esterase
CC on the surface of muscle fibers of the neuromuscular junction.

```

```

CC -!- SUBUNIT: Binds to laminin.
CC -!- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular junction.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Comment=Additional isoforms seem to exist. Isoforms differ in their acetylcholine receptor clustering activity;
CC Name=1;
CC IsoId=P25304-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P25304-2; Sequence=VSP_001365;
CC Name=3;
CC IsoId=P25304-3; Sequence=VSP_001366;
CC Name=4;
CC IsoId=P25304-4; Sequence=VSP_001367;
CC Name=5;
CC IsoId=P25304-5; Sequence=VSP_001368;
CC -!- TISSUE SPECIFICITY: Embryonic nervous system and muscle.
CC -!- DEVELOPMENTAL STAGE: More abundant early in development.
CC -!- PTM: Contains heparan sulfate chains as well as N-linked and O-linked oligosaccharides (By similarity).
CC -!- SIMILARITY: Contains 9 Kazal-like domains.
CC -!- SIMILARITY: Contains 2 laminin EGF-like domains.
CC -!- SIMILARITY: Contains 4 EGF-like domains.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -!- SIMILARITY: Contains 3 laminin G-like domains.
CC -!- CAUTION: It is uncertain whether Met-1, Met-18 or Met-24 is the initiator.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M64780; AAA40703.1; -;
DR EMBL; M64780; AAA40702.1; ALT_INIT.
DR EMBL; S44194; AAB23236.1; -;
DR PIR; JH0399; AGRT.
DR HSSP; P00740; 1EDM.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR003645; FOIN.
DR InterPro; IPR002350; kazal.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR01791; Laminin_G.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF00050; kazal; 9.
DR Pfam; PF00053; laminin_EGF; 2.
DR Pfam; PF00054; laminin_G; 3.
DR Pfam; PF01390; SEA; 1.
DR PRINTS; PRC0011; EGFLAMININ.
DR SMART; SM00180; EGF_Iam; 2.
DR SMART; SM00274; FOLN; 8.
DR SMART; SM00280; KAZAL; 9.
DR SMART; SM0282; IamG; 3.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; 6.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00026; EGF_3; 4.
DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR PROSITE; PS50024; SEA; 1.
KW Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;
KW Laminin EGF-like domain; Proteoglycan; Heparan sulfate.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 1959 AGRIN.
FT DOMAIN 65 137 KAZAL-LIKE 1.
FT DOMAIN 141 212 KAZAL-LIKE 2.
FT DOMAIN 213 284 KAZAL-LIKE 3.

```

FT	DOMAIN	287	356	KAZAL-LIKE 4.	Matches 94; Conservative 28; Mismatches 108; Indels 31; Gaps 10;	
FT	DOMAIN	361	429	KAZAL-LIKE 5.		
FT	DOMAIN	430	494	KAZAL-LIKE 6.		
FT	DOMAIN	495	559	KAZAL-LIKE 7.		
FT	DOMAIN	563	645	KAZAL-LIKE 8.		
FT	DOMAIN	688	741	LAMININ EGF-LIKE 1.		
FT	DOMAIN	742	788	LAMININ EGF-LIKE 2.		
FT	DOMAIN	794	864	SEA.		
FT	DOMAIN	1023	1145	EGF-LIKE 1.		
FT	DOMAIN	1220	1258	LAMININ G-LIKE 1.		
FT	DOMAIN	1263	1439	EGF-LIKE 2.		
FT	DOMAIN	1440	1477	EGF-LIKE 3.		
FT	DOMAIN	1479	1516	LAMININ G-LIKE 2.		
FT	DOMAIN	1526	1708	EGF-LIKE 4.		
FT	DOMAIN	1709	1748	LAMININ G-LIKE 3.		
FT	DOMAIN	1784	1956	SER/THR-RICH.		
FT	DOMAIN	869	992	SER/THR-RICH.		
FT	DOMAIN	1147	1215	POTENTIAL.		
FT	DISULFID	97	116	POTENTIAL.		
FT	DISULFID	105	137	POTENTIAL.		
FT	DISULFID	171	191	POTENTIAL.		
FT	DISULFID	180	212	POTENTIAL.		
FT	DISULFID	244	263	POTENTIAL.		
FT	DISULFID	252	284	POTENTIAL.		
FT	DISULFID	316	335	POTENTIAL.		
FT	DISULFID	324	356	POTENTIAL.		
FT	DISULFID	389	408	POTENTIAL.		
FT	DISULFID	397	429	POTENTIAL.		
FT	DISULFID	454	473	POTENTIAL.		
FT	DISULFID	462	494	POTENTIAL.		
FT	DISULFID	518	538	POTENTIAL.		
FT	DISULFID	527	559	POTENTIAL.		
FT	DISULFID	604	624	POTENTIAL.		
FT	DISULFID	613	645	POTENTIAL.		
FT	DISULFID	688	700	BY SIMILARITY.		
FT	DISULFID	690	707	BY SIMILARITY.		
FT	DISULFID	709	718	BY SIMILARITY.		
FT	DISULFID	721	739	BY SIMILARITY.		
FT	DISULFID	742	754	BY SIMILARITY.		
FT	DISULFID	744	761	BY SIMILARITY.		
FT	DISULFID	763	772	BY SIMILARITY.		
FT	DISULFID	775	786	BY SIMILARITY.		
FT	DISULFID	823	843	POTENTIAL.		
FT	DISULFID	832	864	POTENTIAL.		
FT	DISULFID	1224	1235	BY SIMILARITY.		
FT	DISULFID	1229	1246	BY SIMILARITY.		
FT	DISULFID	1248	1257	BY SIMILARITY.		
FT	DISULFID	1444	1455	POTENTIAL.		
FT	DISULFID	1449	1465	POTENTIAL.		
FT	DISULFID	1467	1476	POTENTIAL.		
FT	DISULFID	1483	1494	BY SIMILARITY.		
FT	DISULFID	1488	1504	BY SIMILARITY.		
FT	DISULFID	1506	1515	BY SIMILARITY.		
FT	DISULFID	1713	1727	BY SIMILARITY.		
FT	DISULFID	1721	1736	BY SIMILARITY.		
FT	DISULFID	1738	1747	BY SIMILARITY.		
FT	CARBOHYD	145	145	N-LINKED (GLCNAC. . .)		(POTENTIAL).
FT	CARBOHYD	572	672	N-LINKED (GLCNAC. . .)		(POTENTIAL).
FT	CARBOHYD	827	827	N-LINKED (GLCNAC. . .)		(POTENTIAL).
FT	CARBOHYD	957	957	N-LINKED (GLCNAC. . .)		(POTENTIAL).
FT	VARSPLIC	1144	1152	Missing (in isoform 2).		
FT	VARSPLIC	1780	1798	Missing (in isoform 3).		
FT	VARSPLIC	1788	1798	Missing (in isoform 4).		
FT	VARSPLIC	1780	1787	Missing (in isoform 5).		
FT	VARIANT	314	314	V -> VTECD (IN A VARIANT).		
FT	SEQUENCE	1959 AA;	208645 MW;	7FEFDFAFF89CC31 CRC64;		
SQ	Query Match		21.3%;	Score 333.5;	DB 1;	Length 1959;
	Best Local Similarity		36.0%;	Pred. No. 4.5e-16;		

RESULT 5

AGRI_DISOM	ID	AGRI_DISOM	STANDARD;	PRT;	1328 AA.
AC	Q90404;	DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)	DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Agtrin (Fragment).	GN	AGRN.		
OS	Discopyge ommata (Electric ray).	OC	Eukaryota; Chordata; Craniata; Vertebrata; Chondrichthyes;		
OC	Elaeobranchii; Squalae; Hypnosqualea; Pristiogaster; Batoidea;	OC	Torpediniformes; Narcinoidei; Narcinidae; Discopyge.		
OX	NCBI_taxid=7785;	RN	[1]		
RA	SEQUENCE FROM N.A.	RA	Smith M.A., Magill-Solc C., Rupp F., Yao Y.-M.M., Schilling J.W.,		
RA	Snow P., McMahan U.J.;	RT	"Isolation and characterization of a cDNA that encodes an agrin		
RT	homolog in the marine ray.";	RL	Mol. Cell. Neurosci. 3:406-417(1992).		
CC	-!	FUNCTION: Component of the basal lamina that causes the			
CC	aggregation of acetylcholine receptors and acetylcholine-esterase				
CC	on the surface of muscle fibers of the neuromuscular junction (By				
CC	similarity).				
CC	-!	SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular			
CC	junction (By similarity).				
CC	-!	SIMILARITY: Contains at least 2 laminin EGF-like domains.			
CC	-!	SIMILARITY: Contains at least 3 EGF-like domains.			
CC	-!	SIMILARITY: Contains 1 SEA domain.			
CC	-!	SIMILARITY: Contains 3 laminin G-like domains.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sb-sib.ch).				
CC	EMBL; L01423; AAA49224.1; ;				
DR	PIR; T43060; T43060.				
DR	HSSP; P00740; IEDM.				
DR	InterPro; IPR00152; Asx_hydroxyl_S.				
DR	InterPro; IPR008985; ConA_like_lec_gl.				
DR	InterPro; IPR000742; EGF_2.				
DR	InterPro; IPR006209; EGF_like.				
DR	InterPro; IPR003645; PolN.				
DR	InterPro; IPR002350; kazal.				
DR	InterPro; IPR002049; Laminin_EGF.				

```

DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR000082; SEA_domain.
DR Pfam: PF00008; EGF; 4.
DR Pfam: PF00050; Kazal; 2.
DR Pfam: PF00053; laminin_EGF; 2.
DR Pfam: PF00054; laminin_G; 3.
DR Pfam: PF01390; SEA; 1.
DR PRINTS; PRA0011; EGF_LAMININ.
DR SMART; SMO0180; EGF_Lam; 2.
DR SMART; SMO0274; FOLN; 2.
DR SMART; SMO0280; KAZAL; 2.
DR SMART; SMO0282; LamG; 3.
DR SMART; SMO0200; SEA; 1.
DR PROSITE; PS00101; ASK_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 5.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00026; EGF_3; 4.
DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
DR PROSITE; PS00024; SEA; 1.
KW Glycoprotein; EGF-like domain; Repeat; Laminin EGF-like domain.
FT NON TER 1
FT DOMAIN 79 132 LAMININ EGF-LIKE 1.
FT DOMAIN 133 179 LAMININ EGF-LIKE 2.
FT DOMAIN 411 533 SEA.
FT DOMAIN 608 644 EGF-LIKE 1.
FT DOMAIN 649 825 LAMININ G-LIKE 1.
FT DOMAIN 865 902 EGF-LIKE 2.
FT DOMAIN 914 1096 LAMININ G-LIKE 2.
FT DOMAIN 1097 1135 EGF-LIKE 3.
FT DOMAIN 1146 1324 LAMININ G-LIKE 3.
FT DISULFID 79 91 BY SIMILARITY.
FT DISULFID 81 98 BY SIMILARITY.
FT DISULFID 100 109 BY SIMILARITY.
FT DISULFID 112 130 BY SIMILARITY.
FT DISULFID 133 145 BY SIMILARITY.
FT DISULFID 135 152 BY SIMILARITY.
FT DISULFID 154 163 BY SIMILARITY.
FT DISULFID 166 177 BY SIMILARITY.
FT DISULFID 612 623 BY SIMILARITY.
FT DISULFID 637 632 BY SIMILARITY.
FT DISULFID 634 643 BY SIMILARITY.
FT DISULFID 869 880 BY SIMILARITY.
FT DISULFID 874 890 BY SIMILARITY.
FT DISULFID 892 901 BY SIMILARITY.
FT DISULFID 1101 1114 BY SIMILARITY.
FT DISULFID 1108 1123 BY SIMILARITY.
FT DISULFID 1125 1134 BY SIMILARITY.
FT CARBOHYD 63 63 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 279 279 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 369 369 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1328 AA; 144018 MW; 79DB81CIAF2A71C18 CRC64;

Query Match
Best Local Similarity 20.1%; Score 314.5; DB 1; Length 1328;
Matches 85; Conservative 44; Mismatches 105; Indels 37; Gaps 12;

OY 1 CERQPCRHGATCMPAGYEYFQCLCRDGFKGLCEHEENPCQLRPFLLHGGFCQ-----GT 55
DB 830 CSPNPKCRKRGKCKMEAEEMFHCSEYVGSFPCADKKNPCD-PNPKCSANCMVLPPEGGS 888
OY 56 KCLCHPGFSGPRCCQGGSHGIAESDHLWLEGGSGNDAPQGYGAYFDHDDGLAPFG-HVFSR 114
DB 889 KCECPMREGEELUCER-----VSEAE-----QQQKAFIPERFNLGSLYEMNNGIHTP-- 933
OY 115 SLFPEPTIELEV--RTSTASGLLWQGVYGEAGGQKDFSLGIQDGHVFRYQLGSGE 172
DB 934 -VSDLLQKLSMEVIFLAKDPNGMIFVNGKTD--GRG-DFVSLNLRDGYLFKFKYLKGA 989
OY 173 ARLVSEDPINDGEWHRYTALRGRRGSIQVDEELVGRSP-----GNVAVNAKGSYVIG 228

```

```

Db 990 AVLRSKAFIPLNWNVVTVERNRGLMKINKDELVSSEPSKSKAPHTALNKEAFYVG 1049
OY 229 GAPD---VATLGGFRSSGITGCVKXNVLHLS 256
Db 1050 GAPDFNKPARAAG--IISGFTGAIQKLSLKS 1078

RESULT 6
AGRI_CHICK
ID AGRI_CHICK STANDARD; PRT; 1955 AA.
AC P31696;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Agrin precursor.
GN AGRN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92232297; PubMed=1314620;
RA Ts'ao K.W.K., Ruegg M.A., Escher G., Kroeger S., McMahon U.J.;
RT "cDNA that encodes active agrin. ";
RL Neuron 8:677-689(1992).
RN [2]
RP ALTERNATIVE SPLICING.
RX MEDLINE=92232298; PubMed=1314621;
RA Ruegg M.A., Ts'ao K.W.K., Horton S.E., Kroeger S., Escher G.,
RA Gensch E.M., McMahon U.J.;
RT "The agrin gene codes for a family of basal lamina proteins that
RT differ in function and distribution.";
RL Neuron 8:691-699(1992).
CC -|- FUNCTION: Component of the basal lamina that causes the
CC aggregation of acetylcholine receptors and acetylcholine-esterase
CC on the surface of muscle fibers of the neuromuscular junction.
CC -|- SUBUNIT: Binds to laminin.
CC -|- SUBCELLULAR LOCATION: Synaptic basal lamina at the neuromuscular
CC junction.
CC -|- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3; Isoforms differ in
CC their acetylcholine receptor clustering activity;
CC Name=1;
CC IsoID=P31696-1; Sequence=Displayed;
CC Name=2; Synonyms=Agrin-related protein 1;
CC IsoID=P31696-2; Sequence=VSP_001370;
CC Name=3; Synonyms=Agrin-related protein 2;
CC IsoID=P31696-3; Sequence=VSP_001369, VSP_001370;
CC -|- PTM: Contains heparan sulfate chains as well as N-linked and O-
CC linked oligosaccharides (By similarity).
CC -|- SIMILARITY: Contains 9 Kazal-like domains.
CC -|- SIMILARITY: Contains 2 laminin EGF-like domains.
CC -|- SIMILARITY: Contains 4 EGF-like domains.
CC -|- SIMILARITY: Contains 1 SEA domain.
CC -|- SIMILARITY: Contains 3 laminin G-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M94271; AAA48585.1; -
CC EMBL; M97371; AAA48586.1; -
CC EMBL; M97372; -; NOT_ANNOTATED_CDS.
CC F00591; ACCH.
CC H00740; LEDM.

```

DR	InterPro; IPR004850; Agrin_NTA				
DR	InterPro; IPR000152; Asx_hydroxy1_S				
DR	InterPro; IPR008985; Cona_like Lec_gl				
DR	InterPro; IPR000742; EGF_2				
DR	InterPro; IPR006209; EGF-like				
DR	InterPro; IPR003645; FOLN				
DR	InterPro; IPR002350; Kazal				
DR	InterPro; IPR002049; Laminin_EGF				
DR	InterPro; IPR001791; Laminin_G				
DR	InterPro; IPR000082; SEA_domain				
DR	InterPro; IPR008993; TIMP_like				
DR	Pfam; PF00008; EGF; 4				
DR	Pfam; PF00050; kazal; 9				
DR	Pfam; PF00053; laminin_EGF; 2				
DR	Pfam; PF00054; laminin_G; 3				
DR	Pfam; PF03146; NFA; 1				
DR	Pfam; PF01390; SEA; 1				
DR	PRINTS; PR00011; EGF_LAMININ				
DR	SMART; SM00180; EGF_Lam; 2				
DR	SMART; SM00274; FOLN; 8				
DR	SMART; SM00280; KAZAL; 9				
DR	SMART; SM00282; LamG; 3				
DR	SMART; SM00200; SEA; 1				
DR	PROSITE; PS00010; ASX_HYDROXYL; 1				
DR	PROSITE; PS00022; EGF_1; 6				
DR	PROSITE; PS01186; EGF_2; 1				
DR	PROSITE; PS00026; EGF_3; 4				
DR	PROSITE; PS50025; Lam_G DOMAIN; 3				
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 1				
DR	PROSITE; PS50024; SEA; 1				
KW	Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;				
KW	Laminin EGF-like domain; Proteoglycan; Heparan sulfate.				
FT	SIGNAL	1	38	POTENTIAL.	
FT	CHAIN	39	1955	AGRN.	
FT	DOMAIN	54	126	KAZAL-LIKE 1.	
FT	DOMAIN	130	201	KAZAL-LIKE 2.	
FT	DOMAIN	202	273	KAZAL-LIKE 3.	
FT	DOMAIN	276	344	KAZAL-LIKE 4.	
FT	DOMAIN	350	418	KAZAL-LIKE 5.	
FT	DOMAIN	419	483	KAZAL-LIKE 6.	
FT	DOMAIN	484	548	KAZAL-LIKE 7.	
FT	DOMAIN	551	633	KAZAL-LIKE 8.	
FT	DOMAIN	675	728	LAMININ EGF-LIKE 1.	
FT	DOMAIN	729	775	LAMININ EGF-LIKE 2.	
FT	DOMAIN	781	851	KAZAL-LIKE 9.	
FT	DOMAIN	1026	1148	SEA.	
FT	DOMAIN	1229	1265	EGF-LIKE 1.	
FT	DOMAIN	1270	1445	LAMININ G-LIKE 1.	
FT	DOMAIN	1446	1483	EGF-LIKE 2.	
FT	DOMAIN	1485	1522	EGF-LIKE 3.	
FT	DOMAIN	1532	1713	LAMININ G-LIKE 2.	
FT	DOMAIN	1714	1752	EGF-LIKE 4.	
FT	DOMAIN	1776	1952	LAMININ G-LIKE 3.	
FT	DOMAIN	856	995	SER/THR-RICH.	
FT	DOMAIN	1150	1219	SER/THR-RICH.	
FT	DISULFID	86	105	POTENTIAL.	
FT	DISULFID	94	126	POTENTIAL.	
FT	DISULFID	160	180	POTENTIAL.	
FT	DISULFID	169	201	POTENTIAL.	
FT	DISULFID	233	252	POTENTIAL.	
FT	DISULFID	241	273	POTENTIAL.	
FT	DISULFID	304	323	POTENTIAL.	
FT	DISULFID	312	344	POTENTIAL.	
FT	DISULFID	378	397	POTENTIAL.	
FT	DISULFID	386	418	POTENTIAL.	
FT	DISULFID	443	462	POTENTIAL.	
FT	DISULFID	451	483	POTENTIAL.	
FT	DISULFID	507	527	POTENTIAL.	
FT	DISULFID	516	548	POTENTIAL.	
FT	DISULFID	592	612	POTENTIAL.	
FT	DISULFID	601	633	POTENTIAL.	
FT	DISULFID	675	687	BY SIMILARITY.	
FT	DISULFID	677	694	BY SIMILARITY.	
DR	InterPro; IPR004850; Agrin_NTA				
DR	InterPro; IPR000152; Asx_hydroxy1_S				
DR	InterPro; IPR008985; Cona_like Lec_gl				
DR	InterPro; IPR000742; EGF_2				
DR	InterPro; IPR006209; EGF-like				
DR	InterPro; IPR003645; FOLN				
DR	InterPro; IPR002350; Kazal				
DR	InterPro; IPR002049; Laminin_EGF				
DR	InterPro; IPR001791; Laminin_G				
DR	InterPro; IPR000082; SEA_domain				
DR	InterPro; IPR008993; TIMP_like				
DR	Pfam; PF00008; EGF; 4				
DR	Pfam; PF00050; kazal; 9				
DR	Pfam; PF00053; laminin_EGF; 2				
DR	Pfam; PF00054; laminin_G; 3				
DR	Pfam; PF03146; NFA; 1				
DR	Pfam; PF01390; SEA; 1				
DR	PRINTS; PR00011; EGF_LAMININ				
DR	SMART; SM00180; EGF_Lam; 2				
DR	SMART; SM00274; FOLN; 8				
DR	SMART; SM00280; KAZAL; 9				
DR	SMART; SM00282; LamG; 3				
DR	SMART; SM00200; SEA; 1				
DR	PROSITE; PS00010; ASX_HYDROXYL; 1				
DR	PROSITE; PS00022; EGF_1; 6				
DR	PROSITE; PS01186; EGF_2; 1				
DR	PROSITE; PS00026; EGF_3; 4				
DR	PROSITE; PS50025; Lam_G DOMAIN; 3				
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 1				
DR	PROSITE; PS50024; SEA; 1				
KW	Glycoprotein; EGF-like domain; Repeat; Alternative splicing; Signal;				
KW	Laminin EGF-like domain; Proteoglycan; Heparan sulfate.				
FT	SIGNAL	1	38	POTENTIAL.	
FT	CHAIN	39	1955	AGRN.	
FT	DOMAIN	54	126	KAZAL-LIKE 1.	
FT	DOMAIN	130	201	KAZAL-LIKE 2.	
FT	DOMAIN	202	273	KAZAL-LIKE 3.	
FT	DOMAIN	276	344	KAZAL-LIKE 4.	
FT	DOMAIN	350	418	KAZAL-LIKE 5.	
FT	DOMAIN	419	483	KAZAL-LIKE 6.	
FT	DOMAIN	484	548	KAZAL-LIKE 7.	
FT	DOMAIN	551	633	KAZAL-LIKE 8.	
FT	DOMAIN	675	728	LAMININ EGF-LIKE 1.	
FT	DOMAIN	729	775	LAMININ EGF-LIKE 2.	
FT	DOMAIN	781	851	KAZAL-LIKE 9.	
FT	DOMAIN	1026	1148	SEA.	
FT	DOMAIN	1229	1265	EGF-LIKE 1.	
FT	DOMAIN	1270	1445	LAMININ G-LIKE 1.	
FT	DOMAIN	1446	1483	EGF-LIKE 2.	
FT	DOMAIN	1485	1522	EGF-LIKE 3.	
FT	DOMAIN	1532	1713	LAMININ G-LIKE 2.	
FT	DOMAIN	1714	1752	EGF-LIKE 4.	
FT	DOMAIN	1776	1952	LAMININ G-LIKE 3.	
FT	DOMAIN	856	995	SER/THR-RICH.	
FT	DOMAIN	1150	1219	SER/THR-RICH.	
FT	DISULFID	86	105	POTENTIAL.	
FT	DISULFID	94	126	POTENTIAL.	
FT	DISULFID	160	180	POTENTIAL.	
FT	DISULFID	169	201	POTENTIAL.	
FT	DISULFID	233	252	POTENTIAL.	
FT	DISULFID	241	273	POTENTIAL.	
FT	DISULFID	304	323	POTENTIAL.	
FT	DISULFID	312	344	POTENTIAL.	
FT	DISULFID	378	397	POTENTIAL.	
FT	DISULFID	386	418	POTENTIAL.	
FT	DISULFID	443	462	POTENTIAL.	
FT	DISULFID	451	483	POTENTIAL.	
FT	DISULFID	507	527	POTENTIAL.	
FT	DISULFID	516	548	POTENTIAL.	
FT	DISULFID	592	612	POTENTIAL.	
FT	DISULFID	601	633	POTENTIAL.	
FT	DISULFID	675	687	BY SIMILARITY.	
FT	DISULFID	677	694	BY SIMILARITY.	

Query Match

Best Local Similarity 19.3%; Score 301.5; DB 1; Length 1955;

Mismatches 87; Conservative 38; Mismatches 108; Indels 85; Gaps 13;

Matches 87; Conservative 38; Mismatches 108; Indels 85; Gaps 13;

QY 1 CERQPCQCATCPAGEYEFQCLCRDGFKGLCEHEENPCQLREPCLEHGTC-----OGT 55

Db 1450 CHPNFCHRGASCHVKEAMFHCECLHSLTGTCDARNPCD-PTFCHI.SATCLVLPGGGA 1508

QY 56 RCLCLPGFSGPCQGGHGAIESDWHLEGGNDAPGGYAGYFHDGDFLAPFGHVSRS 115

Db 1509 MCACPMGREGFCEP-----VTEQD-----HT 1530

QY 116 LPEVPE-----TILEVTRSTASGLLWQGVVEBAGGCKPFI 153

Db 1531 MPFLEPFNGFSYLELNGLQTLFLTCRQMSMEVFLAKSPGMIFYNGQKTD--GRG-DFV 1587

QY 154 SLGLQDGHLVFRYOLGSGEARLVSDPDINDGEWHRVATLREGRGSIQVDGBELVGRSP 213

Db 1588 SLLAHGYLEVRYDLGKGAVALRKEFPVLNWTISVLLERGRKGVMMINGERVWGESP 1647

QY 214 G-----PNVAVAKGSVYIGGAPDVATLT--GGRFSGGITCVKNIHLVHSAKPGAPPQPLD 268

Db 1648 KSRKVPFAFLNKPEFVYGGAPDFSKLAAAAIAIATSFYGAIVQISII-----KGVP---LLK 1700

QY 269 LOH-RAQAGANT---RPC 282

Db 1701 EQHIRSAVEISITFRAHPC 1718

RESULT 7

FAT_DROME STANDARD; PRT; 5147 AA.

AC P3350; OSVQXS;

DT 01-FEB-1994 (Rel. 28, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cadherin-related tumor suppressor precursor (fat protein).

GN FT OR CG3352.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92069752; PubMed=1959133;
RA Mahoney P.A., Weber U., Onofrechuk P., Biessmann H., Bryant P.J.,
RD Goodman C.S.;
RT "The fat tumor suppressor gene in Drosophila encodes a novel member
RL of the cadherin gene superfamily. ";
RN Call 67.853-868(1991).
RP [2]
RC SEQUENCE FROM N.A.
RD STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pabloe B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flossler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy K., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirezka R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Waberman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster. ";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: Could function as a cell-adhesion protein. Acts as a
CC tumor suppressor. Required for correct morphogenesis.
CC -!- SIMILARITY: Contains 34 cadherin domains.
CC -!- SIMILARITY: Contains 5 EGF-like domains.
CC -!- SIMILARITY: Contains 2 laminin G-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M80537; AAA28530.1; -;
CC EMBL; AE003577; AAF51036.1; -;
CC HSSP; P00740; LEDM.
CC FlyBase; Fgn0001075; ft.
CC GO; GO:0005887; C:integral to plasma membrane; NAS.
CC GO; GO:0008014; F:calcium-dependent cell-cell adhesion; NAS.
CC GO; GO:0016339; P:calcium-dependent cell-cell adhesion; NAS.
CC GO; GO:0008283; P:cell proliferation; IMP.

DR GO; GO:0000904; P:cellular morphogenesis during differentiation; IMP.
DR GO; GO:0045317; P:equator specification; IMP.
DR GO; GO:0004598; P:establishment of epithelial cell polarity; IMP.
DR GO; GO:0007446; P:imaginal disc growth; IMP.
DR GO; GO:0018149; P:protein-protein cross-linking; IPI.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR008985; ConA_like_lect_gl.
DR InterPro; IPR000742; EGF_2
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR006210; EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00028; cadherin; 34.
DR Pfam; PF00054; EGF; 4.
DR Pfam; PF00054; Laminin_G; 2.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 34.
DR SMART; SM00181; EGF; 4.
DR SMART; SM0282; Lamg; 2.
DR PROSITE; PS00232; CADHERIN_1; 22.
DR PROSITE; PS00268; CADHERIN_2; 34.
DR PROSITE; PS00022; EGF_1; 4.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS50036; EGF_3; 4.
DR PROSITE; PS50035; LAM_G_DOMAIN; 2.
KW Cell adhesion; Signal; Transmembrane; Glycoprotein; Calcium-binding;
KW Repeat; EGF-like domain.
FT SIGNAL 1 35
FT CHAIN 36 5147
FT DOMAIN 36 4583
FT TRANSMEM 4584 4609
FT DOMAIN 4610 5147
FT DOMAIN 36 156
FT DOMAIN 157 270
FT DOMAIN 271 382
FT DOMAIN 383 494
FT DOMAIN 495 599
FT DOMAIN 600 708
FT DOMAIN 709 820
FT DOMAIN 821 942
FT DOMAIN 943 1049
FT DOMAIN 1050 1153
FT DOMAIN 1154 1278
FT DOMAIN 1279 1384
FT DOMAIN 1385 1489
FT DOMAIN 1490 1601
FT DOMAIN 1602 1713
FT DOMAIN 1714 1823
FT DOMAIN 1824 1922
FT DOMAIN 1923 2027
FT DOMAIN 2028 2167
FT DOMAIN 2168 2278
FT DOMAIN 2279 2385
FT DOMAIN 2386 2491
FT DOMAIN 2492 2596
FT DOMAIN 2597 2703
FT DOMAIN 2704 2810
FT DOMAIN 2811 2913
FT DOMAIN 2914 3013
FT DOMAIN 3014 3124
FT DOMAIN 3125 3229
FT DOMAIN 3230 3334
FT DOMAIN 3335 3439
FT DOMAIN 3440 3545
FT DOMAIN 3546 3651
FT DOMAIN 3652 3755
FT DOMAIN 3950 4011
FT DOMAIN 4013 4049
FT DOMAIN 4052 4090
FT DOMAIN 4092 4128
FT DOMAIN 4129 4320
FT DOMAIN 4321 4362
FT DOMAIN 4402 4569
FT DISULFID 3954 3966
POTENTIAL.
CADHERIN-RELATED TUMOR SUPPRESSOR.
EXTRACELLULAR (POTENTIAL).
POTENTIAL.
CYTOPLASMIC (POTENTIAL).
CADHERIN 1.
CADHERIN 2.
CADHERIN 3.
CADHERIN 4.
CADHERIN 5.
CADHERIN 6.
CADHERIN 7.
CADHERIN 8.
CADHERIN 9.
CADHERIN 10.
CADHERIN 11.
CADHERIN 12.
CADHERIN 13.
CADHERIN 14.
CADHERIN 15.
CADHERIN 16.
CADHERIN 17.
CADHERIN 18.
CADHERIN 19.
CADHERIN 20.
CADHERIN 21.
CADHERIN 22.
CADHERIN 23.
CADHERIN 24.
CADHERIN 25.
CADHERIN 26.
CADHERIN 27.
CADHERIN 28.
CADHERIN 29.
CADHERIN 30.
CADHERIN 31.
CADHERIN 32.
CADHERIN 33.
CADHERIN 34.
EGF-LIKE 1.
EGF-LIKE 2.
EGF-LIKE 3.
EGF-LIKE 4.
LAMININ G-LIKE 1.
EGF-LIKE 5.
LAMININ G-LIKE 2.
BY SIMILARITY.

QY	Db	Accession	Score	DB 1	DB 2	Length	DB 1	DB 2
				17.28	Score 269	DB 1	Length 5147	
				29.7%	Pred. No. 5.1e-11			
				94	Conservative	39	Mismatches 102	Indels 82
								Gaps 19
QY	1	CERPCOHGATCW--PAGEVEFQCLCRDGFKGLCEHEENPCQLRPCLHGTC	4113	---	OG 54			
Db	4056	CYSKPCRNKGGSCQSPDGSSYF-CLCRPFRNQCBSVDSR-PNFCPLHGGLCVSLKFG	4113					
QY	55	TRCLCLFGFGPRCQQGGHGAESAHWLRSGGNDAPGOYGAYFDDGFLAPPGHVFSR	114					
Db	4114	YKNCCTEGRYGRHCERFS-----YG--FQPLSVMTFF-----	4143					
QY	115	SLEVPFTELEVRTSTASGLLWQ-GVEVGEAGQKDFSLGLQDCHLVFVYQLGSGEA	173					
Db	4144	ALDVTNTDSDIVFATFPNLSLNLYGM---QSGRRSDFLAELVHGRAYF-----SSGGA	4196					
QY	174	R-----LVSDEPNDGSHWRTALREGRGSIQV-----DGE--ELVGRSP-----	214					
Db	4197	RTAISVTIAGNLDAGGWHKVTATENGVMNLSVAKCADSGDVCTECLPDSSCYADVYG	4256					
QY	215	PNVAVN-AKGSVIYG---APDVAITLFGFRFSSGITGVKXNVLHSGARQPAPPQPLDI-	269					
Db	4257	PVGTLPNFKPLMIGLSSADPILFRPGVHSDDLVGC-----LHSVHIGG---RALNLS	4308					
QY	270	---OHRQAQANTRPC 282						
Db	4309	LFLQKQKILACNRQAC 4325						

RESULT 8	SLT2_HUMAN	ID	SLT2_HUMAN	STANDARD	PRT	1529 AA
AC	O94813	O95710	O95710	O95710		
DT	10-OCT-2003	(Rel. 42, Created)				
DT	10-OCT-2003	(Rel. 42, Last sequence update)				
DT	10-OCT-2003	(Rel. 42, Last annotation update)				
DE	Slit homolog 2 protein precursor (h-Slit-2).					
GN	SLIT2 OR SLIT2					
OS	Homo sapiens (Human)					
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
OC	Mammalia; Euteheria; Primates; Catarrhini; Homnidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.					
RC	TISSUE=Fetal lung;					
RX	MEDLINE=99033071; PubMed=9813312;					
RA	Itoh A., Miyabayashi T., Ohno M., Sakano S.;					
RT	"Cloning and expressions of three mammalian homologues of Drosophila					
RT	slit suggest possible roles for slit in the formation and maintenance					
RT	of the nervous system";					
RL	Brain Res. Mol. Brain Res. 62:175-186(1998).					
RN	[2]					
RP	SEQUENCE FROM N.A. (ISOFORM 3), AND TISSUE SPECIFICITY.					
RC	TISSUE=Fetal brain, and Fetal kidney;					
RX	MEDLINE=99279238; PubMed=10349621;					
RA	Holmes G.P., Negus K., Burridge L., Raman S., Algar B., Yamada T.,					
RA	Little M.H.;					
RT	"Distinct but overlapping expression patterns of two vertebrate slit					
RT	homologs implies functional roles in CNS development and					
RT	organogenesis.";					
RL	Mech. Dev. 79:57-72(1998).					
RN	[3]					
RP	SEQUENCE FROM N.A. (ISOFORM 2), SEQUENCE OF 1122-1129, FUNCTION,					
RC	SUBUNIT, AND SUBCELLULAR LOCATION.					
RX	MEDLINE=99200391; PubMed=10102268;					
RA	Brose K., Bland K.S., Wang K.H., Arnott D., Henzel W., Goodman C.S.,					
RA	Tessier-Lavigne M., Kidd T.;					
RT	"Slit proteins bind Robo receptors and have an evolutionarily					
RT	conserved role in repulsive axon guidance.";					
RL	Cell 96:795-806(1999).					
CC	-!- FUNCTION: Plays a role in axon guidance as a repulsive ligand for					
CC	Robo receptors preventing inappropriate midline crossing.					
CC	-!- SUBUNIT: Binds robo proteins with high affinity.					
CC	-!- SUBCELLULAR LOCATION: Secreted. The C-terminal cleavage protein is					
CC	more diffusible than the larger N-terminal protein that is more					
CC	tightly cell associated.					
CC	-!- ALTERNATIVE PRODUCTS:					
CC	Event-Alternative splicing; Named isoforms=3;					
CC	Name=1;					
CC	IsoId=O94813-1; Sequence=Displayed;					
CC	Name=2;					
CC	IsoId=O94813-2; Sequence=VSP_050035, VSP_050036;					
CC	Name=3;					
CC	IsoId=O94813-3; Sequence=VSP_050036;					
CC	-!- TISSUE SPECIFICITY: Fetal lung and kidney, and adult spinal cord.					
CC	Weak expression in adult adrenal gland, thyroid, trachea and other					
CC	tissues examined.					
CC	-!- SIMILARITY: Contains 7 EGF-like domains.					
CC	-!- SIMILARITY: Contains 23 leucine-rich (LRR) repeats.					
CC	-!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.					
CC	-!- SIMILARITY: Contains 1 laminin G-like domain.					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.ebi.ac.uk/announcements/					
CC	or send an email to license@ebi.ac.uk).					
DR	EMBL; AB017168; BAA35185.1;					

QY 92 -----PGQYGA-----YFHDGDFLAFPGHVFSLPEVPEF-TELEVRISTAS 133
 Db 1141 NEPICQCLFFGQGEKCKELVSNFINKESYLQIP-----SAKVRPQTNILOQTATDEDS 1194
 QY 134 GLLMQGVEVGEAGQGGDFISLQDGHVFRYQLGSGEARLV-SEDPINDGQEWHRVIAL 192
 Db 1195 GILLYKG-----DKDHIHAVELYRGRVRSYDTGSHPSAISVSVETINDGNFHVIVELL 1246
 QY 193 REGRRGSLQVDSG--BELVSGRSPGNVAVNAKGSVYIGGAP---DVATL---TGRFRSSGI 245
 Db 1247 ALDQSLSLVSDGNGPKIITNLS---KQSTLNFPLSDPLVYGGMPGKSNVSLRQAFQNGQTSF 1304
 QY 246 TGCVKNLVLS 256
 Db 1305 HGCIIRNLVINS 1315

RESULT 9
 FAT2 DROME STANDARD; PRT: 4705 AA.
 AC O9VW71; O9VSS1;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Putative fat-like cadherin-related tumor suppressor homolog precursor.
 DE FAT2 OR CG7749.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RX STRAIN=Berkeley;
 RC MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.C., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P.,
 RA Burkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry M., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K.I., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirkas R., Tector K., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."
 Science 287:2185-2195 (2000).
 RL [2]
 RN REVISIONS.
 RX MEDLINE=24246069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celnik S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review."
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [3]
 RP SEQUENCE OF 3837-4705 FROM N.A.
 RC STRAIN=Berkeley; TISSUE=Ovary;
 RX MEDLINE=24246066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celnik S.E.,
 RT "A drosophila full-length cDNA resource."
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -!- SIMILARITY: Contains 34 cadherin domains.
 CC -!- SIMILARITY: Contains 5 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 laminin G-like domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE003515; AAF49078.2; -
 DR EMBL; AY060955; AAL28503.1; ALT_INIT.
 DR EMBL; AY118666; AAMS0035.1; ALT_INIT.
 DR HSSP; P15116; INCI.
 DR FlyBase; FBgn0036930; fat2
 DR GO; GO:0005887; C:integral to plasma membrane; ISS.
 DR GO; GO:0016339; F:calcium-dependent cell adhesion molecule ac. . ; ISS.
 DR InterPro; IPR00152; Asx hydroxyl S.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR008985; ConA like_lec_gl.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00028; cadherin; 31.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF00054; laminin G; 1.
 DR PRINTS; PRO0205; CADHERIN.
 DR SMART; SM00112; CA; 34.
 DR SMART; SM00181; EGF; 6.
 DR SMART; SM00282; LamG; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00232; CADHERIN_1; 18.
 DR PROSITE; PS50268; CADHERIN_2; 34.
 DR PROSITE; PS00022; EGF_1; 5.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS50026; EGF_3; 5.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 1.
 DR Hypothetical protein; Cell adhesion; Signal; Glycoprotein;
 DR Transmembrane; Calcium; Calcium-binding; Repeat; EGF-like domain.
 FT SIGNAL 1 35
 FT CHAIN 36 4705 PUTATIVE FAT-LIKE CADHERIN-RELATED TUMOR
 FT DOMAIN 36 1647 SUPPRESSOR HOMOLOG.
 FT TRANSMEM 1648 1668 EXTRACELLULAR (POTENTIAL).
 FT

FT DOMAIN	1669	4705	Matches	59;	Conservative	43;	Mismatches	88;	Indels	56;	Gaps	10;
FT DOMAIN	60	180	22	CLCRGFKGDLCEHEN	-----	PCOLRREPCLHGTCQGRCLCLPFGSPRQOQSGHGI	76					
FT DOMAIN	181	288	3867	CYCRFGDKKCKETVNACTDPCSPQRICMPSGALGYQCYCPKFGSTYCKERKSKCS	3926							
FT DOMAIN	285	397	77	ABS-DWHL--	EGSGNDAPGOYGVAFH	-----	DDGFLAFPHGVFSRSLPEVP	120				
FT DOMAIN	398	504	3927	NESCDMGLFTAVSFGK	-----	SYAHKINKVAKTLENGF	-----	3963				
FT DOMAIN	505	610	121	ETIELEVRTSTASGLLLMGVVEGRAGQGDIFISLGLQDGHVLFVRYQLGSGEARL	-VSED	179						
FT DOMAIN	611	713	3964	-SYSLQIRTVQQTGTLVYASGKV	-----	DYNILEIINGAVQYRFDLGSSEGVISVSSI	4015					
FT DOMAIN	770	874	180	PINDGEHVEVTLREGRRGSIQVDBEELVSGRSPGNVAVNAKGS	-VYIGG	-APDVATL	236					
FT DOMAIN	875	977	4016	NISDGEWHQISLERSLNSAKVMVDNKHVSHGSAFGVNGILNIQSDNDFVGAERVHPHSII	4075							
FT DOMAIN	978	1085	237	TGGRFSSGITGCVKML	252							
FT DOMAIN	1086	1195	4076	GYEDIQRFIGCWANI	4091							
FT DOMAIN	1191	1296										
FT DOMAIN	1297	1402										
FT DOMAIN	1403	1503										
FT DOMAIN	1504	1609										
FT DOMAIN	1610	1714										
FT DOMAIN	1715	1812										
FT DOMAIN	1813	1929										
FT DOMAIN	1930	2030										
FT DOMAIN	2031	2137										
FT DOMAIN	2138	2238										
FT DOMAIN	2239	2338										
FT DOMAIN	2339	2465										
FT DOMAIN	2466	2567										
FT DOMAIN	2568	2670										
FT DOMAIN	2671	2779										
FT DOMAIN	2780	2876										
FT DOMAIN	2877	2983										
FT DOMAIN	2984	3088										
FT DOMAIN	3089	3185										
FT DOMAIN	3186	3289										
FT DOMAIN	3290	3394										
FT DOMAIN	3395	3499										
FT DOMAIN	3500	3604										
FT DOMAIN	3605	3712										
FT DOMAIN	3819	3879										
FT DOMAIN	3881	3919										
FT DOMAIN	3937	4121										
FT DOMAIN	4129	4166										
FT DOMAIN	4168	4205										
FT DOMAIN	4243	4279										
FT DISULFID	3823	3835										
FT DISULFID	3830	3867										
FT DISULFID	3859	3878										
FT DISULFID	3885	3896										
FT DISULFID	3890	3907										
FT DISULFID	3909	3918										
FT DISULFID	4133	4144										
FT DISULFID	4138	4154										
FT DISULFID	4156	4165										
FT DISULFID	4172	4183										
FT DISULFID	4177	4193										
FT DISULFID	4195	4204										
FT DISULFID	4247	4258										
FT DISULFID	4252	4267										
FT DISULFID	4269	4278										
FT CARBOHYD	65	65										
FT CARBOHYD	156	156										
FT CARBOHYD	364	364										
FT CARBOHYD	779	779										
FT CARBOHYD	843	843										
FT CARBOHYD	923	923										
FT CARBOHYD	1106	1106										
FT CARBOHYD	1198	1198										
FT CARBOHYD	1312	1312										
FT CARBOHYD	1439	1439										
FT CARBOHYD	1473	1473										
FT CARBOHYD	1511	1511										
FT CONFLICT	3962	3962										
SQ SEQUENCE	4705 AA;	524564 MW;	6D387A489D2C33DE	CRC64;								

Best Local Similarity 15.5%; Score 242; DB 1; Length 4705; Pred. No. 3.9e-09; 27.0%;

RESULT 10

STAN_DROME ID STAN_DROME STANDARD; PRT; 3579 AA.
 AC Q8V5N8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Protocadherin-like wing polarity protein stan precursor (Starry night protein) (Flamingo protein).
 GN STAN OR FMI OR CG11895.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N. A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
 RC TISSUE=Embryo;
 RX MEDLINE=20025940; PubMed=10556066;
 RA Chae J.W., Kim M.-J., Goo J.H., Collier S., Gubb D., Charlton J., Adler P.N., Park W.J.,
 RT "The Drosophila tissue polarity gene starry night encodes a member of the protocadherin family."
 RL Development 126:5421-5429(1999).
 RN [2]
 RP SEQUENCE FROM N. A., FUNCTION, AND SUBCELLULAR LOCATION.
 RX MEDLINE=99418630; PubMed=10490098;
 RA Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L., Takeichi M., Uemura T.;
 RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell polarity under the control of frizzled."
 RL Cell 98:585-595(1999).
 RN [3]
 RP SEQUENCE FROM N. A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton R.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abril J.P., Agayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu D., Beasley E.M., Beeson K.Y., Benos P.V., Bernick B.P., Bhandari D., Bolshakov S., Borokova D., Botchan M.R., Bouck J., Brockton P., Brottier P., Burtis K.C., Busam D.A., Butler H., Cadien E., Center A., Chandra I., Cherry J.M., Cusumano S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harlow N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houshon K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Paclib J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamas I., Simpson M., Stupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "the genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [4]
 RN REVISIONS.
 RP MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminler J.S., Millburn G.H., Prochuk S.B.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Battencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 CC -1- FUNCTION: Involved in the fz signaling pathway that controls wing
 CC tissue polarity. Also mediates homophilic cell adhesion. May play
 CC a role in initiating prehair morphogenesis. May play a critical
 CC role in tissue polarity and in formation of normal dendrite
 CC fields.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: In the pupal wing, expressed at relatively
 CC even levels in all regions. Abundant in 6-9 hour embryos.
 CC -1- EXPRESSED AT HIGHER LEVELS IN PUPAE THAN LARVAE.
 CC -1- DEVELOPMENTAL STAGE: At 12 hours after puparium formation (apf),
 CC is concentrated at proximal and distal cell boundaries with little
 CC or no expression at anterior and posterior boundaries. When
 CC preairs emerge at 30-36 hours apf, expression becomes evenly
 CC distributed again along the whole cell boundary.
 CC -1- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
 CC -1- SIMILARITY: Contains 8 cadherin domains.
 CC -1- SIMILARITY: Contains 4 EGF-like domains.
 CC -1- SIMILARITY: Contains 2 laminin G-like domains.
 CC -1- SIMILARITY: Contains 1 GPS domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF172329; AAF02618.1; -
 CC EMBL; AB028498; BAA84069.1; -
 CC EMBL; AE003828; AAF58763.3; -
 CC HSSP; P08709; I8F9; -
 CC FlyBase; FBgn0024836; stan.
 CC GO; GO:0016021; C:integral to membrane; NAS.

DR GO; GO:008084; F:calcium-dependent cell adhesion molecule ac.; IMP.
 DR GO; GO:005057; F:receptor signaling protein activity; IMP.
 DR GO; GO:0016358; P:dendrite morphogenesis; IEP.
 DR GO; GO:007222; P:fertilized signaling pathway; IMP.
 DR GO; GO:0007367; P:segment polarity determination; IMP.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR008985; ConA-like_lectin.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR000832; GPCR_secretin.
 DR InterPro; IPR001879; hormn_receptor.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR000203; PKD_cys_fich.
 DR Pfam; PF00002; 7tm_2; 1.
 DR Pfam; PF00028; cadherin; 8.
 DR Pfam; PF00008; EGF; 3.
 DR Pfam; PF01825; GPS; 1.
 DR Pfam; PF02793; HRM; 1.
 DR Pfam; PF00053; laminin_EGF; 1.
 DR Pfam; PF00054; laminin_G; 1.
 DR PRINTS; PR00205; CADHERIN.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR PRINTS; PR00249; GPCR_SECRETIN.
 DR SMART; SM00112; CA; 8.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00303; GPS_1.
 DR SMART; SM00282; LamG; 2.
 DR PROSITE; PS00232; CADHERIN_1; 6.
 DR PROSITE; PS0268; CADHERIN_2; 8.
 DR PROSITE; PS00022; EGF_1; 4.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS00026; EGF_3; 3.
 DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
 DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
 DR PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
 DR PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
 DR PROSITE; PS50221; GPS; 1.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 2.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 DR Cell adhesion: Developmental protein; G-protein coupled receptor;
 KW Calcium-binding; Repeat; Signal; Transmembrane; EGF-like domain;
 KW Laminin EGF-like domain; Glycoprotein.
 KW SIGNAL 1 29
 KW CHAIN 30 3579
 FT PROTOCADHERIN-LIKE WING POLARITY PROTEIN
 FT STAN.
 FT DOMAIN 30 2816 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 2817 2837 1 (POTENTIAL).
 FT DOMAIN 2838 2845 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 2846 2866 2 (POTENTIAL).
 FT DOMAIN 2867 2883 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 2884 2904 3 (POTENTIAL).
 FT DOMAIN 2905 2919 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 2920 2940 4 (POTENTIAL).
 FT DOMAIN 2941 2959 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 2960 2980 5 (POTENTIAL).
 FT DOMAIN 2981 3000 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 3001 3021 6 (POTENTIAL).
 FT DOMAIN 3022 3031 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 3032 3052 7 (POTENTIAL).
 FT DOMAIN 3053 3079 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 360 464 CADHERIN 1.
 FT DOMAIN 465 581 CADHERIN 2.
 FT DOMAIN 582 689 CADHERIN 3.
 FT DOMAIN 690 794 CADHERIN 4.
 FT DOMAIN 795 897 CADHERIN 5.
 FT DOMAIN 898 1007 CADHERIN 6.
 FT DOMAIN 1008 1113 CADHERIN 7.
 FT DOMAIN 1114 1220 CADHERIN 8.
 FT DOMAIN 1482 1518 EGF-LIKE 1 CALCIUM-BINDING.
 FT DOMAIN 1556 1753 LAMININ G-LIKE 1.
 FT DOMAIN 1756 1922 EGF-LIKE 2 CALCIUM-BINDING.

FT DOMAIN 1796 1963 LAMININ G-LIKE 2.
 FT DOMAIN 1965 2000 EGF-LIKE 3 CALCIUM-BINDING.
 FT DOMAIN 2091 2126 EGF-LIKE 4 CALCIUM-BINDING.
 FT DOMAIN 2744 2802 GFS.
 FT DOMAIN 140 143 POLY-HIS.
 FT DOMAIN 155 159 POLY-ARG.
 Query Match 14.4%; Score 225; DB 1; Length 3579;
 Best Local Similarity 22.4%; Pred. No. 4.7e-08;
 Matches 83; Conservativity 52; Mismatches 120; Indels 116; Gaps 17;
 QY 1 CERQCOHGATCMP-----AGEY-----EFQCLCRDGFKGD-----LCEH 35
 Db 1422 CVREPCLNFEELTVLKFNGASEFTHSDVLPFRPIYPVNTFACSCPEGFTGSKHYLCDT 1481
 QY 36 ENPFCOLREPCVHGCTC-----QCTRCICLPFSRPRCOQSGH-----GIAESDWHLEG 85
 Db 1482 EVDLC-YSDFCQNGTGVRRREGGYTCVCFSTHTGQCETGVGHLRPPCPSETCEGSLCLS 1540
 QY 86 SGGNDAPGOYGA-----YFDDGFLAPPG-----HVFSRSLPEVPEPIELEVRTSTASG 134
 Db 1541 NYFSSQPPPYTATCELRARAFGRNSFLTFESLQKQRHFN-----LKLRFATVQENG 1591
 QY 135 LLLWQGVVEAGQKDKTISLQDCHLVFRYQLGSGEARL--VSEDPINDGEWHRYTAL 192
 Db 1592 LLLYN---GRYNELHDFIALEIHEGHVSVFSLGDHSEISVJQEAQVSDGKWHQVEV 1647
 QY 193 RECRGSIQVDCGE---LVSGR-----SPGNVAVNAKGSVYI 227
 Db 1648 YLNRSVTLVLDNCDYTAIALSGQLGDRWSCANRTTLKDKRCSLLTETCHRELDLTGFLQV 1707
 QY 228 GGADPVAT---ITGGRFSSGITGVKNLVL-----HSRPGAPPPOPLDLOH 271
 Db 1708 GGLPRIPAFPPVTRDF-----VGCISDLRIDDFVLDLNSYVADNGTLGACQPKAPL----- 1759
 QY 272 RAQAGANTRPC 282
 Db 1760 -----CQSEPC 1765
 RESULT 11
 ID LMA2 HUMAN STANDARD; PRT; 3110 AA.
 AC P2403; Q14736; Q931022;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Laminin alpha-2 chain precursor (Laminin M chain) (Merosin heavy chain).
 GN LAMA2 OR LAMM.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=94124633; PubMed=8294519;
 RA Vuolteenaho R., Nissinen M., Sainio K., Byers M., Eddy R., Hirvonen H., Shows T.B., Sariola H., Engvall E., Tryggvason K.; "Human laminin M chain (merosin): complete primary structure, chromosomal assignment, and expression of the M and A chain in human fetal tissues."; J. Cell Biol. 124:381-394(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97066955; PubMed=8910357;
 RA Zhang X., Vuolteenaho R., Tryggvason K.; "Structure of the human laminin alpha2-chain gene (LAMA2), which is affected in congenital muscular dystrophy."; J. Biol. Chem. 271:27664-27669(1996).
 RN [3]
 RP SEQUENCE OF 1981-3110 FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Placenta;
 RX MEDLINE=90238994; PubMed=2185464;
 RA Ehrig K., Leivo I., Argraves W.S., Ruoslahti E., Engvall E.; "Merosin, a tissue-specific basement membrane protein, is a laminin-like protein."; Proc. Natl. Acad. Sci. U.S.A. 87:3264-3268(1990).
 RN [4]
 RP VARIANTS GLN-545; HIS-619; LEU-919; HIS-2586 AND LYS-2614.
 RA Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z., Marzluf G.A., Amato A.A., Mendell J.R.; "Novel single base polymorphisms and rare sequence variants in the laminin 2-chain coding region detected by RNA/SSCP analysis."; Hum. Mutat. 13:174-174(1999).
 RN [5]
 RP ERRATUM.
 RA Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z., Marzluf G.A., Amato A.A., Mendell J.R.; Hum. Mutat. 13:340-340(1999).
 RL [6]
 RN VARIANT MDC1A PRO-2564.
 RP MEDLINE=21476011; PubMed=11591858;
 RX He Y., Jones K.J., Vignier N., Morgan G., Chevally M., Barois A., Estourmet-Mathiaud B., Hori H., Mizuta T., Tome F.M.S., North K.N., Guicheney P.; "Congenital muscular dystrophy with primary partial laminin alpha-2 chain deficiency: molecular study."; Neurology 57:1319-1322(2001).
 RL [7]
 RN VARIANTS MDC1A TYR-527 AND ARG-862.
 RP MEDLINE=22439669; PubMed=12552556;
 RX Tezak Z., Prandini P., Boscaro M., Marin A., Devaney J., Marino M., Fanin M., Trevisan C.P., Park J., Tyson W., Finkel R., Garcia C., Angelini C., Hoffman E.P., Pegoraro E.; "Clinical and molecular study in congenital muscular dystrophy with partial laminin alpha-2 (LAMA2) deficiency."; Hum. Mutat. 21:1103-111(2003).
 RT -0- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.
 CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end. The alpha-2 chain is a subunit of laminin-2 (merosin) and laminin-4 (S-merosin).
 CC -1- SUBCELLULAR LOCATION: Extracellular; found in the basement membranes (major component).
 CC -1- TISSUE SPECIFICITY: Placenta, striated muscle, peripheral nerve, cardiac muscle, pancreas, lung, spleen, kidney, adrenal gland, skin, testis, meninges, choroid plexus, and some other regions of the brain; not in liver, thymus and bone.
 CC -1- DOMAIN: The alpha-helical domains I and II are thought to interact with other laminin chains to form a coiled coil structure.
 CC -1- DOMAIN: Domains VI, IV and G are globular.
 CC -1- DISEASE: Defects in LAMA2 are the cause of merosin-deficient congenital muscular dystrophy type IA (MDC1A) [MIM:607855]. MDC1A is characterized by difficulty walking, hypotonia, proximal weakness, hyporeflexia, and white matter hypodensity on MRI.
 CC -1- SIMILARITY: Contains 1 laminin N-terminal domain.
 CC -1- SIMILARITY: Contains 17 laminin EGF-like domains.
 CC -1- SIMILARITY: Contains 5 laminin G-like domains.
 CC -1- SIMILARITY: Contains 5 laminin G-like domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z26653; CAA81394.1; -

DR	EMBL; U66796; AAB19388.1; --	DR	InterPro; IPR008979; Gal_bind_like
DR	EMBL; U66733; AAB19388.1; JOINED	DR	InterPro; IPR000034; Laminin_B
DR	EMBL; U66734; AAB19388.1; JOINED	DR	InterPro; IPR002049; Laminin_EGF
DR	EMBL; U66735; AAB19388.1; JOINED	DR	InterPro; IPR001791; Laminin_G
DR	EMBL; U66736; AAB19388.1; JOINED	DR	InterPro; IPR008211; LamNT
DR	EMBL; U66737; AAB19388.1; JOINED	DR	Pfam; PF00052; laminin_B_2
DR	EMBL; U66738; AAB19388.1; JOINED	DR	Pfam; PF00053; laminin_EGF; 14
DR	EMBL; U66739; AAB19388.1; JOINED	DR	Pfam; PF00054; laminin_G; 5
DR	EMBL; U66740; AAB19388.1; JOINED	DR	Pfam; PF00055; laminin_Nterm; 1
DR	EMBL; U66741; AAB19388.1; JOINED	DR	PRINTS; PR00011; EGF_LAMININ
DR	EMBL; U66742; AAB19388.1; JOINED	DR	ProDom; PD003031; Laminin_B; 1
DR	EMBL; U66743; AAB19388.1; JOINED	DR	SMART; SM00180; EGF_Lam; I5
DR	EMBL; U66745; AAB19388.1; JOINED	DR	SMART; SM00281; LamE; 2
DR	EMBL; U66746; AAB19388.1; JOINED	DR	SMART; SM00282; LamG; 5
DR	EMBL; U66747; AAB19388.1; JOINED	DR	SMART; SM00136; LamNT; 1
DR	EMBL; U66748; AAB19388.1; JOINED	DR	PROSITE; PS00022; EGF_1; 11
DR	EMBL; U66749; AAB19388.1; JOINED	DR	PROSITE; PS01186; EGF_2; 3
DR	EMBL; U66750; AAB19388.1; JOINED	DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 14
DR	EMBL; U66751; AAB19388.1; JOINED	DR	PROSITE; PS50025; LAM_G_DOMAIN; 5
DR	EMBL; U66752; AAB19388.1; JOINED	KW	Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
DR	EMBL; U66753; AAB19388.1; JOINED	KW	Laminin EGF-like domain; Cell adhesion; Repeat; Signal; Polymorphism;
DR	EMBL; U66754; AAB19388.1; JOINED	FT	Signal
DR	EMBL; U66755; AAB19388.1; JOINED		
DR	EMBL; U66756; AAB19388.1; JOINED		
DR	EMBL; U66757; AAB19388.1; JOINED		
DR	EMBL; U66758; AAB19388.1; JOINED		
DR	EMBL; U66759; AAB19388.1; JOINED		
DR	EMBL; U66760; AAB19388.1; JOINED		
DR	EMBL; U66761; AAB19388.1; JOINED		
DR	EMBL; U66762; AAB19388.1; JOINED		
DR	EMBL; U66763; AAB19388.1; JOINED		
DR	EMBL; U66764; AAB19388.1; JOINED		
DR	EMBL; U66765; AAB19388.1; JOINED		
DR	EMBL; U66766; AAB19388.1; JOINED		
DR	EMBL; U66768; AAB19388.1; JOINED		
DR	EMBL; U66769; AAB19388.1; JOINED		
DR	EMBL; U66770; AAB19388.1; JOINED		
DR	EMBL; U66771; AAB19388.1; JOINED		
DR	EMBL; U66772; AAB19388.1; JOINED		
DR	EMBL; U66773; AAB19388.1; JOINED		
DR	EMBL; U66774; AAB19388.1; JOINED		
DR	EMBL; U66775; AAB19388.1; JOINED		
DR	EMBL; U66776; AAB19388.1; JOINED		
DR	EMBL; U66777; AAB19388.1; JOINED		
DR	EMBL; U66778; AAB19388.1; JOINED		
DR	EMBL; U66779; AAB19388.1; JOINED		
DR	EMBL; U66780; AAB19388.1; JOINED		
DR	EMBL; U66781; AAB19388.1; JOINED		
DR	EMBL; U66782; AAB19388.1; JOINED		
DR	EMBL; U66783; AAB19388.1; JOINED		
DR	EMBL; U66784; AAB19388.1; JOINED		
DR	EMBL; U66785; AAB19388.1; JOINED		
DR	EMBL; U66786; AAB19388.1; JOINED		
DR	EMBL; U66787; AAB19388.1; JOINED		
DR	EMBL; U66788; AAB19388.1; JOINED		
DR	EMBL; U66789; AAB19388.1; JOINED		
DR	EMBL; U66790; AAB19388.1; JOINED		
DR	EMBL; U66791; AAB19388.1; JOINED		
DR	EMBL; U66792; AAB19388.1; JOINED		
DR	EMBL; U66793; AAB19388.1; JOINED		
DR	EMBL; U66794; AAB19388.1; JOINED		
DR	EMBL; U66795; AAB19388.1; JOINED		
DR	EMBL; MS9832; ARA63215.1; --		
DR	PIR; FX0082; MHUMH		
DR	HSSP; Q60675; 1QU0		
DR	Genew; HGNC:6482; LAMA2		
DR	MIM; 156225; --		
DR	MIM; 607855; --		
DR	GO; GO:0005604; C:basement membrane; TAS		
DR	GO; GO:0005198; F:structural molecule activity; TAS		
DR	GO; GO:0007517; P:muscle development; TAS		
DR	InterPro; IPR008985; Cons_like Lec_Gl		
DR	InterPro; IPR006209; EGF_like		

DR	InterPro; IPR008979; Gal_bind_like		
DR	InterPro; IPR000034; Laminin_B		
DR	InterPro; IPR002049; Laminin_EGF		
DR	InterPro; IPR001791; Laminin_G		
DR	InterPro; IPR008211; LamNT		
DR	Pfam; PF00052; laminin_B_2		
DR	Pfam; PF00053; laminin_EGF; 14		
DR	Pfam; PF00054; laminin_G; 5		
DR	Pfam; PF00055; laminin_Nterm; 1		
DR	PRINTS; PR00011; EGF_LAMININ		
DR	ProDom; PD003031; Laminin_B; 1		
DR	SMART; SM00180; EGF_Lam; I5		
DR	SMART; SM00281; LamE; 2		
DR	SMART; SM00282; LamG; 5		
DR	SMART; SM00136; LamNT; 1		
DR	PROSITE; PS00022; EGF_1; 11		
DR	PROSITE; PS01186; EGF_2; 3		
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 14		
DR	PROSITE; PS50025; LAM_G_DOMAIN; 5		
KW	Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;		
KW	Laminin EGF-like domain; Cell adhesion; Repeat; Signal; Polymorphism;		
FT	Signal		
QY	122 TIELEVRTRSTAGLLMGQVEVGERAGQKDFISLQDQHCHLVFRYQLGSGEARLVSQEDPI 181	Query Match	14.3%; Score 223.5; DB 1; Length 3110;
Db	2788 TIELEVRTEAESGLLFYMA----AINHADPATVQLRNGLFFYSYDLGSGDPTHTMIPTKI 2842	Best Local Similarity	36.4%; Pred. No. 5.1e-08;
QY	182 NDEGWHRTVALREGRRGSIQVDEBELVSRGSPGNVA--VNAKGSVIYGGAP-DVATILTG 238	Matches	59; Conservative 22; Mismatches 68; Indels 13; Gaps 5;
Db	2843 NDCQWHKIKIMRSKQEGILYVDG---ASNRITSPKADILDVWGMIVYVGLPVIYTRRI 2899		
QY	239 GRPSSGITGVKLVHSAAPGAPPDPDLQHRQAQANTR 280		
Db	2900 GPTVYSDIGCVRN--LHMAEAPADLEQPTSSFVHGTCFANAQ 2939		

RESULT 12

NK3A_RAT
ID NX3A_RAT STANDARD; PRT; 1578 AA.
AC Q07310; Q07280; Q07311; Q07312; Q07313; Q07314;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neurexin 3-alpha precursor (Neurexin III-alpha).
GN NRXN3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Scurionathii; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]_TaxID=10116;
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RA MEDLINE=93342001; PubMed=8341647;
RX Ushkaryov Y.A., Suedhof T.C.;
RT "Neurexin III alpha: extensive alternative splicing generates
RT membrane-bound and soluble forms";
RL Proc. Natl. Acad. Sci. U.S.A. 90:6410-6414(1993).
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=95209856; PubMed=7695896;
RA Ulrich B., Ushkaryov Y.A., Suedhof T.C.;
RT "Cartography of neurexins: more than 1000 isoforms generated by
RT alternative splicing and expressed in distinct subsets of neurons";
RN Neuron 14:497-507(1995).
RN [3]
RP SEQUENCE OF N-TERMINUS, AND INTERACTION WITH NEUREXOPHILIN 1.
RX MEDLINE=99074239; PubMed=9856994;

RA Missler M., Hammer R.E., Suedhof T.C.;
 RT "Neurexophilin binding to alpha-neurexins. A single LNS domain
 RT functions as an independently folding ligand-binding unit."
 RL J. Biol. Chem. 273:34716-34723(1998).
 RN [4]
 RP INTERACTION WITH CASK.
 RX MEDLINE=96256685; PubMed=8786425;
 RA Hata Y., Butz S., Suedhof T.C.;
 RT "CASK: a novel dig/PSD95 homolog with an N-terminal
 RT calmodulin-dependent protein kinase domain identified by interaction
 RT with neurexins."
 RL J. Neurosci. 16:2488-2494(1996).
 RN [5]
 RP INTERACTION WITH ALPHA-DYSTROGLYCAN.
 RX MEDLINE=21363578; PubMed=11470830;
 RA Sugita S., Saito P., Tang J., Campbell K., Suedhof T.C.;
 RT "A stoichiometric complex of neurexins and dystroglycan in brain."
 RL J. Cell Biol. 154:1435-1445(2001).
 CC -!- FUNCTION: Neuronal cell surface protein that may be involved in
 CC cell recognition and cell adhesion. May mediate intracellular
 CC signaling.
 CC -!- SUBUNIT: The laminin G-like domain 2 binds to neurexophilin 1.
 CC Isoforms alpha 4B bind to alpha-dystroglycan. The cytoplasmic C-
 CC terminal region binds to CASK.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential). Some
 CC isoforms seems to be secreted.
 CC -!- ALTERNATIVE PRODUCTS.
 CC Event=Alternative splicing; Named isoforms=1;
 CC Comment=At least 288 isoforms may be produced by alternative
 CC splicing. There is a combination of five alternative spliced
 CC domains at sites 1, 3, 4 and 5, each consisting of modular
 CC sequences (A-L) that seem to be used independently. Isoforms
 CC Alpha 5A-type to Alpha 5H-type lack the transmembrane domain;
 CC Name=Alpha 1A3A4A5I.
 CC IsoId=Q07310-1; Sequence=Displayed;
 CC -!- TISSUE SPECIFICITY: Brain.
 CC -!- SIMILARITY: Contains 6 laminin G-like domains.
 CC -!- SIMILARITY: Contains 3 EGF-like domains.
 CC -!- SIMILARITY: Belongs to the neurexin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L14851; AAA02857.1; - -
 DR EMBL; L14851; AAA02856.1; - -
 DR EMBL; L14851; AAA02858.1; - -
 DR EMBL; L14851; AAA02853.1; - -
 DR EMBL; L14851; AAA02855.1; - -
 DR EMBL; L14851; AAA02854.1; - -
 DR HSSP; Q63373; 1C4F.
 DR InterPro; IPR000152; Ask_hydroxyl_S.
 DR InterPro; IPR008985; ConA_like Lec_gl.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR003585; Neurexin-like.
 DR Pfam; PF00008; EGF_3.
 DR Pfam; PF00054; laminin_G; 6.
 DR SMART; SM00294; 4.lim; 1.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00282; LamG; 6.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; FALSE_NEG.
 DR PROSITE; PS50026; EGF_3; 3.
 DR PROSITE; PS50025; Lam_G_DOMAIN; 6.
 DR Signal; P550025; Lam_G_DOMAIN; 6.
 KW Signal; Transmembrane; Repeat; Cell adhesion; EGF-like domain;
 KW Glycoprotein; Alternative splicing.

FT	SIGNAL	1	27		
FT	CHAIN	28	1578		NEUREXIN 3-ALPHA.
FT	DOMAIN	28	1503		EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1504	1524		POTENTIAL.
FT	DOMAIN	1525	1578		CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	28	202		LAMININ G-LIKE 1.
FT	DOMAIN	198	235		EGF-LIKE 1.
FT	DOMAIN	260	444		LAMININ G-LIKE 2.
FT	DOMAIN	451	643		LAMININ G-LIKE 3.
FT	DOMAIN	647	684		EGF-LIKE 2.
FT	DOMAIN	689	861		LAMININ G-LIKE 4.
FT	DOMAIN	875	1050		LAMININ G-LIKE 5.
FT	DOMAIN	1053	1090		EGF-LIKE 3.
FT	DOMAIN	1094	1294		LAMININ G-LIKE 6.
FT	CARBOHYD	58	58		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	105	105		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	761	761		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1193	1193		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1291	1291		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1335	1335		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1500	1500		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DOMAIN	282	282		POLY-SER.
FT	DOMAIN	1323	1326		POLY-THR.
FT	DOMAIN	1329	1332		POLY-THR.
FT	DOMAIN	1407	1410		POLY-THR.
FT	DOMAIN	1422	1425		POLY-SER.
FT	DOMAIN	1511	1514		POLY-ALA.
FT	VARSPPLIC	237	242		Missing (in some isoforms Alpha 1C and Alpha 1D). /FTId=VSP_003521.
FT	VARSPPLIC	237	256		Missing (in some isoforms Alpha 1F). /FTId=VSP_003522.
FT	VARSPPLIC	243	256		Missing (in some isoforms Alpha 1E). /FTId=VSP_003523.
FT	VARSPPLIC	253	256		Missing (in some isoforms Alpha 1B and Alpha 1D). /FTId=VSP_003524.
FT	VARSPPLIC	754	763		DCIRLNCNS\$ -> G (in some isoforms Alpha 3B). /FTId=VSP_003525.
FT	VARSPPLIC	1210	1239		Missing (in some isoforms Alpha 4B). /FTId=VSP_003526.
FT	VARSPPLIC	1369	1371		Missing (in some isoforms Alpha 5B, Alpha 5D, Alpha 5F, Alpha 5H, Alpha 5J and Alpha 5L). /FTId=VSP_003527.
FT	VARSPPLIC	1372	1478		Missing (in some isoforms Alpha 5A, Alpha 5B, Alpha 5C, Alpha 5D, Alpha 5E, Alpha 5F, Alpha 5G, Alpha 5H, Alpha 5K and Alpha 5L). /FTId=VSP_003528.
FT	VARSPPLIC	1479	1578		ANPTEGIRRVPGASEVIRESNSTGMVYGVAAAALCILII LYAMYKYRNRDEGSYQVDETRNYISNSAQSNGLMKEKQA SSKSGHKQKNDKEYYV -> ARSNAARITPCRPYMDMA TLLHLYPSHLHLCLIDTLPFPFPFPPLSLALIKF MCCHPPP (in some isoforms Alpha 5A and Alpha 5B). /FTId=VSP_003529.
FT	VARSPPLIC	1479	1578		ANPTEGIRRVPGASEVIRESNSTGMVYGVAAAALCILII LYAMYKYRNRDEGSYQVDETRNYISNSAQSNGLMKEKQA SSKSGHKQKNDKEYYV -> VLERRIILNKTNHPKSL OSKTC (in some isoforms Alpha 5C and Alpha 5D). /FTId=VSP_003530.
FT	VARSPPLIC	1479	1578		ANPTEGIRRVPGASEVIRESNSTGMVYGVAAAALCILII LYAMYKYRNRDEGSYQVDETRNYISNSAQSNGLMKEKQA SSKSGHKQKNDKEYYV -> DILLKSF (in some isoforms Alpha 5E and Alpha 5F). /FTId=VSP_003531.
FT	VARSPPLIC	1479	1578		ANPTEGIRRVPGASEVIRESNSTGMVYGVAAAALCILII LYAMYKYRNRDEGSYQVDETRNYISNSAQSNGLMKEKQA SSKSGHKQKNDKEYYV -> ATTTTKSNQECGNSICP RAPLHNFLI (in some isoforms Alpha 5G and Alpha 5H).

Alpha 5H
/FTid=VSP_003532
SEQUENCE 1578 AA; 173993 MW; 936CF8529143D0C7 CRC64;
Query Match 13.7% ; Score 214; DB 1; Length 1578;
Best Local Similarity 26.1% ; Pred. No. 1.2e-07;
Matches 73; Conservative 38; Mismatches 121; Indels 48; Gaps 10;
Oy 27 GPKGDLCE-----HEENPCQLRECLHGVC-----CGTRCLC-LPGF 63
Db 171 GFKGLMIDLKGNSEFRLLSQSVLRAGPCGER-PCENGGICFLLDGHTCDCTGTG 229
Oy 64 SGPRCQSGHGIAESDWHLEGGSDNAPQCYGAYFDDGLFAPPGHVFERSLPEVPEII 123
Db 230 GGTLCCSEDSVQPCGLSHLWSEQGRKAREENVATFRGSYLSV--DLQNPVQQSSSEI 287
Oy 124 ELEVRTASGLLWQGVVEGAEQCKDIFSLGLDGHVFRYQLSGEARLIVSDPI-- 181
Db 288 TLSFPTWQRNGLL-----HTGKSADYVNLALKDGNVSLVNLGSGAPEAIVE-PVNG 339
Oy 182 --NDGWHRYTALREGRRGSIQVDEELVSRGPGPNVAVNNAKGVVYIGAPDVALTGG 239
Db 340 KENDNAMHDVKNVNTLNKQVIVSVDGLITITTYTQEDYTMLGSDSSYVGFSPSTADLPS 399
Oy 240 RFSSGITGVKLVL--HSRARPAGAPPQPLDLQHRAQAGA 277
Db 400 PVSNNFMGLCKEYVYKNNDIR-----LESLRLARIGA 431
RESULT 13
CLR2_HUMAN
ID _CLR2_HUMAN STANDARD; PRT; 2923 AA.
AC_Q9R2C4; Q92566;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cadherin EGF LAG seven-type receptor 2 precursor (Epidermal
DE growth factor-like 2) (Multiple epidermal growth factor-like domains
DE 3) (Flamingo 1).
GN_CELSR2 OR CDHF10 OR EGF2 OR MEGF3 OR KIAA0279.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RX_MEDLINE=20363102; PubMed=10907856;
RA_Vincent J.B., Skaug J., Scherer S.W.;
RT "The human homologue of flamingo, EGF2, encodes a brain-expressed
RT large cadherin-like protein with epidermal growth factor-like domains,
RT and maps to chromosome 1p13.3-p11.1";
RL_DNA Res. 7:233-235(2000).
[2]
RP SEQUENCE FROM N.A.
RA_Suwa M., Sato T., Okouchi I., Arita M., Futami K., Matsumoto S.,
RA_Tsutsumi S., Aburatani H., Asai K., Akiyama Y.;
RT "Genome-wide discovery and analysis of human seven transmembrane helix
RT receptor genes";
RL_Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE OF 516-2923 FROM N.A.
RC_TISSUE=Brain;
RX_MEDLINE=97191544; PubMed=9039502;
RA_Nagase T., Seki N., Iahikawa K.-I., Ohira M., Kawabayasi Y.,
RA_Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT analysis of cDNA clones from cell line KG-1 and brain";
RT_DNA Res. 3:321-329(1996).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Highest expression in brain and testis.

-!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
-!- SIMILARITY: Contains 9 cadherin domains.
-!- SIMILARITY: Contains 8 EGF-like domains.
-!- SIMILARITY: Contains 2 laminin G-like domains.
-!- SIMILARITY: Contains 1 GPS domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
EMBL AF234887; AAC00080.1; --
EMBL AB065955; BAC06169.1; --
EMBL DB7469; BA13407.1; --
HSP: P15116; INCU.
GeneW; HGNC:3231; CELSR2.
MIM: 604265; --
GO: GO:0016021; C: integral to membrane; NAS.
GO: GO:004930; F: G-protein coupled receptor activity; NAS.
GO: GO:0007186; P: G-protein coupled receptor protein signalin.; NAS.
InterPro: IPR001152; Asx_hydroxy1_S.
InterPro: IPR002126; Cadherin.
InterPro: IPR008985; ConA_like_lec_gl.
InterPro: IPR00742; EGF 2.
InterPro: IPR001881; EGF_Ca.
InterPro: IPR006209; EGF_like.
InterPro: IPR000832; GPCR_secretin.
InterPro: IPR001879; hormn_receptor.
InterPro: IPR002049; Laminin_EGF.
InterPro: IPR001791; Laminin_G.
InterPro: IPR002203; PKD_cys_rich.
Pfam: PF00002; 7tm_2_1_cys_rich.
Pfam: PF00028; cadherin; 8.
Pfam: PF00008; EGF; 5.
Pfam: PF01825; GPS; 1.
Pfam: PF02793; HRM; 1.
Pfam: PF00053; laminin_EGF; 1.
Pfam: PF00054; laminin_G; 1.
PRINTS: PR00205; CADHERIN.
PRINTS: PR00011; EGF_LAMININ.
PRINTS: PR00249; GPCRSECRETIN.
SMART: SM00112; CA; 9.
SMART: SM00179; EGF_CA; 1.
SMART: SM00303; GPS; 1.
SMART: SM00008; Horm; 1.
SMART: SM00282; LamG; 2.
PROSITE: PS00010; ASX_HYDROXYL; 2.
PROSITE: PS00232; CADHERIN_1; 7.
PROSITE: PS0268; CADHERIN_2; 9.
PROSITE: PS00022; EGF_1; 6.
PROSITE: PS01186; EGF_2; 4.
PROSITE: PS00026; EGF_3; 6.
PROSITE: PS00649; G_PROTEIN_REC_P2_1; FALSE_NEG.
PROSITE: PS00650; G_PROTEIN_REC_P2_2; FALSE_NEG.
PROSITE: PS0227; G_PROTEIN_REC_P2_3; 1.
PROSITE: PS0261; G_PROTEIN_REC_P2_4; 1.
PROSITE: PS0221; GPS; 1.
PROSITE: PS0025; LAM_G_DOMAIN; 2.
PROSITE: PS01248; LAMININ_TYPE_EGF; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;
KW Developmental protein; Hydroxylation; Signal.
FT SIGNAL 1 31
FT CHAIN 32 2923 CADHERIN EGF LAG SEVEN-PASS G-TYPE
FT RECEPTOR 2.
FT DOMAIN 32 2380 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2381 2401 1 (POTENTIAL).
FT DOMAIN 2402 2416 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 2417 2437 2 (POTENTIAL).
FT DOMAIN 2438 2438 EXTRACELLULAR (POTENTIAL).

FT	TRANSMEM	2439	2459	3 (POTENTIAL).	FT	CARBHYD	2345	2345	N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	DOMAIN	2460	2480	CYTOPLASMIC (POTENTIAL).	SO	SEQUENCE	2923 AA;	317447 MW;	382757D315158ED8 CRC64;			
FT	TRANSMEM	2481	2501	4 (POTENTIAL).								
FT	DOMAIN	2502	2519	EXTRACELLULAR (POTENTIAL).								
FT	TRANSMEM	2520	2540	5 (POTENTIAL).								
FT	DOMAIN	2541	2560	CYTOPLASMIC (POTENTIAL).								
FT	TRANSMEM	2561	2581	6 (POTENTIAL).								
FT	DOMAIN	2582	2591	EXTRACELLULAR (POTENTIAL).								
FT	TRANSMEM	2592	2612	7 (POTENTIAL).								
FT	DOMAIN	2613	2623	CYTOPLASMIC (POTENTIAL).								
FT	DOMAIN	182	289	CADHERIN 1.	QY	1	CRPQCQHGATCMPAGEYE	-----FCCLCRDGFKGLDLCHEENP	39			
FT	DOMAIN	290	399	CADHERIN 2.	Db	1232	CLREFCENTMRCVSLREFDSSAFFIASSVLPFRPHVYVGLRCRCPPGFTGDYCEYEDL	-----	1291			
FT	DOMAIN	400	505	CADHERIN 3.	QY	40	COLRBPCLHGTCQ	---GTRCLCLPFGSPRCQGG	---SGH	---GIASRDWH	---LEGS	86
FT	DOMAIN	506	610	CADHERIN 4.	Db	1292	CYSR-PCGPHGRCSRREGYTCLCRDGYTGERCEVASRGRCTPGVKNGGTCVNLVGG	-----	1350			
FT	DOMAIN	611	712	CADHERIN 5.	QY	87	GGNAP-GQYGAYFHDDGFLAPPGHVF	---SRSLPE	---VPETIELEVRTSTAGLLMQGVE	142		
FT	DOMAIN	713	815	CADHERIN 6.	Db	1351	FKDCPCSGDFEKPYCVQVITRSPFAHSFITFRGLRFRHFTLALSFAIKERDGLLLYN	-----	1407			
FT	DOMAIN	816	921	CADHERIN 7.	QY	143	VEAGQKDFISLGLQDGHVFRYQVGSSEARLVSEDP	---INDGWRHV	-----	189		
FT	DOMAIN	922	1023	CADHERIN 8.	Db	1408	GRFNEKDFVALEVIQEQVLTFFSAGSTTTVPFVGGVSDGQWHTVQLKYNKPLLG	-----	1466			
FT	DOMAIN	1028	1146	CADHERIN 9.	QY	190	TALREG	-----RRGSIQVDGELVSRSPGNVAVNAKSVYIG	228			
FT	DOMAIN	1228	1286	EGF-LIKE 1, CALCIUM-BINDING.	Db	1467	QTGLPQSPSEORVAVVTVVGGCDTGVALRFGVSLGNYSCAAQGTGGSKSLDLTGLPLLG	-----	1526			
FT	DOMAIN	1288	1324	EGF-LIKE 2, CALCIUM-BINDING.	QY	229	GAPDVATLTGGRFSSGIGTCVKNLVLSAR	-----PGAP	262			
FT	DOMAIN	1328	1366	EGF-LIKE 3, CALCIUM-BINDING.	Db	1527	GVFDLPESFPVRNQ-FVCGMRNLQVDSRHDMADFIANNQVTPGCP	-----	1572			
FT	DOMAIN	1367	1571	LAMININ G-LIKE 1.								
FT	DOMAIN	1574	1610	EGF-LIKE 4, CALCIUM-BINDING.								
FT	DOMAIN	1614	1791	LAMININ G-LIKE 2.								
FT	DOMAIN	1793	1828	EGF-LIKE 5, CALCIUM-BINDING.								
FT	DOMAIN	1829	1867	EGF-LIKE 6, CALCIUM-BINDING.								
FT	DOMAIN	1883	1922	EGF-LIKE 7, CALCIUM-BINDING.								
FT	DOMAIN	1923	1955	EGF-LIKE 8, CALCIUM-BINDING.								
FT	DOMAIN	2316	2368	GPS.								
FT	DOMAIN	2743	2752	POLY-GLU.								
FT	DISULFID	1232	1243	BY SIMILARITY.								
FT	DISULFID	1237	1274	BY SIMILARITY.								
FT	DISULFID	1276	1285	BY SIMILARITY.								
FT	DISULFID	1292	1303	BY SIMILARITY.								
FT	DISULFID	1297	1312	BY SIMILARITY.								
FT	DISULFID	1314	1323	BY SIMILARITY.								
FT	DISULFID	1332	1343	BY SIMILARITY.								
FT	DISULFID	1337	1353	BY SIMILARITY.								
FT	DISULFID	1355	1365	BY SIMILARITY.								
FT	DISULFID	1578	1589	BY SIMILARITY.								
FT	DISULFID	1583	1598	BY SIMILARITY.								
FT	DISULFID	1600	1609	BY SIMILARITY.								
FT	DISULFID	1757	1808	BY SIMILARITY.								
FT	DISULFID	1802	1817	BY SIMILARITY.								
FT	DISULFID	1819	1828	BY SIMILARITY.								
FT	DISULFID	1832	1843	BY SIMILARITY.								
FT	DISULFID	1837	1855	BY SIMILARITY.								
FT	DISULFID	1857	1866	BY SIMILARITY.								
FT	DISULFID	1887	1899	BY SIMILARITY.								
FT	DISULFID	1889	1906	BY SIMILARITY.								
FT	DISULFID	1908	1921	BY SIMILARITY.								
FT	DISULFID	1924	1936	BY SIMILARITY.								
FT	DISULFID	1926	1943	BY SIMILARITY.								
FT	DISULFID	1945	1954	BY SIMILARITY.								
FT	MOD_RES	1591	1591	HYDROXYLATION (POTENTIAL).								
FT	MOD_RES	1810	1810	HYDROXYLATION (POTENTIAL).								
FT	CARBHYD	486	486	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	557	557	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	701	701	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	1036	1036	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	1076	1076	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	1182	1182	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	1212	1212	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	1501	1501	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	1565	1565	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	1741	1741	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	1827	1827	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	1900	1900	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	2024	2024	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	2043	2043	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	2061	2061	N-LINKED (GLCNAC. . .) (POTENTIAL).								
FT	CARBHYD	2323	2323	N-LINKED (GLCNAC. . .) (POTENTIAL).								

RESULT 14
 CLR2 RAT
 ID CLR2 RAT STANDARD; PRT; 2144 AA.
 AC Q9QIF2;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cadherin EGF LAG seven-pass G-type receptor 2 (Multiple epidermal growth factor-like domains 3) (Fragment).
 GN CELSR2 OR MEGF3
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=98360089, PubMed=9893030;
 RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
 RT EGF-like motifs by motif-trap screening.";
 RL Genomics 51:27-34(1998).
 CC -!- FUNCTION: Receptor that may have an important role in cell/cell signaling during nervous system formation.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: Expressed in the brain. High expression in cerebellum and olfactory bulb. Weaker expression in cerebral cortex, hippocampus and brain stem.
 CC -!- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
 CC -!- SIMILARITY: Contains 4 cadherin domains.
 CC -!- SIMILARITY: Contains 8 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 laminin G-like domains.
 CC -!- SIMILARITY: Contains 1 laminin EGF-like domain.
 CC -!- SIMILARITY: Contains 1 GPS domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@sib.ch).

CC CC EMBL; AB011529; BA88687.1; .

DR DR HSSP; P00740; IEDM.

DR DR GO; GO:0016021; C:Integral to membrane; ISS.

DR DR GO; GO:0004930; F:G-protein coupled receptor activity; ISS.

DR DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . ; ISS.

DR DR InterPro; IPR000152; Asx_hydroxyl_S.

DR DR InterPro; IPR002126; Cadherin.

DR DR InterPro; IPR008985; Cona_like Lec_gl.

DR DR InterPro; IPR000742; EGF_2.

DR DR InterPro; IPR001881; EGF_Ca.

DR DR InterPro; IPR006209; EGF like.

DR DR InterPro; IPR000832; GPCR secretin.

DR DR InterPro; IPR001879; hctm receptor.

DR DR InterPro; IPR002049; Laminin_EGF.

DR DR InterPro; IPR001791; Laminin_G.

DR DR InterPro; IPR002029; PKD_cys_rich.

DR DR Pfam; PF00002; 7tm_2; 1.

DR DR Pfam; PF00028; cadherin; 3.

DR DR Pfam; PF00008; EGF; 5.

DR DR Pfam; PF01825; GFS; 1.

DR DR Pfam; PF02793; HRM; 1.

DR DR Pfam; PF00053; laminin_EGF; 1.

DR DR Pfam; PF00054; laminin_G; 1.

DR DR PRINTS; PR00205; CADHERIN.

DR DR PRINTS; PR00011; EGF_LAMININ.

DR DR PRINTS; PR00249; GPCRSECRETIN.

DR DR SMART; SM00112; CA; 3.

DR DR SMART; SM00179; EGF_CA; 1.

DR DR SMART; SM00303; GFS; 1.

DR DR SMART; SM00008; Hctm; 1.

DR DR SMART; SM00282; Lamg; 2.

DR DR PROSITE; PS00010; ASX_HYDROXYL; 2.

DR DR PROSITE; PS00232; CADHERIN_1; 3.

DR DR PROSITE; PS50268; CADHERIN_2; 4.

DR DR PROSITE; PS00022; EGF_1; 6.

DR DR PROSITE; PS01186; EGF_2; 4.

DR DR PROSITE; PS50026; EGF_3; 6.

DR DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.

DR DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.

DR DR PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.

DR DR PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.

DR DR PROSITE; PS50221; GFS; 1.

DR DR PROSITE; PS50025; LAM_G_DOMAIN; 2.

DR DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.

DR DR G-protein coupled receptor; Transmembrane; Glycoprotein;

DR DR EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;

DR DR Developmental protein; Hydroxylation.

FT NON_TER 1

FT DOMAIN 1 1605 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1606 1626 1 (POTENTIAL).

FT DOMAIN 1627 1641 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1642 1662 2 (POTENTIAL).

FT DOMAIN 1663 1663 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1664 1684 3 (POTENTIAL).

FT DOMAIN 1685 1705 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1706 1726 4 (POTENTIAL).

FT DOMAIN 1727 1744 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1745 1765 5 (POTENTIAL).

FT DOMAIN 1766 1789 6 (POTENTIAL).

FT TRANSMEM 1790 1810 7 (POTENTIAL).

FT DOMAIN 1811 1816 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1817 1837 7 (POTENTIAL).

FT DOMAIN 1838 2144 CYTOPLASMIC (POTENTIAL).

FT CADHERIN 1.

FT DOMAIN 41 40 CADHERIN 2.

FT DOMAIN 147 248 CADHERIN 3.

FT DOMAIN 253 371 CADHERIN 4.

FT DOMAIN 453 511 EGF-LIKE 1. CALCIUM-BINDING.

FT DOMAIN 513 549 EGF-LIKE 2. CALCIUM-BINDING.

FT DOMAIN 553 591 EGF-LIKE 3. CALCIUM-BINDING.

FT DOMAIN 592 796 LAMININ G-LIKE 1.

799 835 EGF-LIKE 4, CALCIUM-BINDING.

1016 1016 LAMININ G-LIKE 2.

1018 1053 EGF-LIKE 5, CALCIUM-BINDING.

1054 1092 EGF-LIKE 6, CALCIUM-BINDING.

1108 1147 EGF-LIKE 7, CALCIUM-BINDING.

1148 1180 EGF-LIKE 8, CALCIUM-BINDING.

1541 1593 GFS.

1968 1973 POLY-GLU.

457 468 BY SIMILARITY.

469 499 BY SIMILARITY.

501 510 BY SIMILARITY.

517 528 BY SIMILARITY.

522 537 BY SIMILARITY.

539 548 BY SIMILARITY.

557 568 BY SIMILARITY.

562 578 BY SIMILARITY.

580 590 BY SIMILARITY.

803 814 BY SIMILARITY.

823 834 BY SIMILARITY.

1022 1033 BY SIMILARITY.

1027 1042 BY SIMILARITY.

1044 1053 BY SIMILARITY.

1057 1068 BY SIMILARITY.

1062 1080 BY SIMILARITY.

1082 1091 BY SIMILARITY.

1112 1124 BY SIMILARITY.

1114 1131 BY SIMILARITY.

1133 1146 BY SIMILARITY.

1149 1161 BY SIMILARITY.

1151 1168 BY SIMILARITY.

1170 1179 BY SIMILARITY.

816 816 HYDROXYLATION (POTENTIAL).

1035 1035 HYDROXYLATION (POTENTIAL).

261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).

301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).

407 407 N-LINKED (GLCNAC. . .) (POTENTIAL).

437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).

726 726 N-LINKED (GLCNAC. . .) (POTENTIAL).

790 790 N-LINKED (GLCNAC. . .) (POTENTIAL).

966 966 N-LINKED (GLCNAC. . .) (POTENTIAL).

1052 1052 N-LINKED (GLCNAC. . .) (POTENTIAL).

1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).

1249 1249 N-LINKED (GLCNAC. . .) (POTENTIAL).

1268 1268 N-LINKED (GLCNAC. . .) (POTENTIAL).

1286 1286 N-LINKED (GLCNAC. . .) (POTENTIAL).

1548 1548 N-LINKED (GLCNAC. . .) (POTENTIAL).

1570 1570 N-LINKED (GLCNAC. . .) (POTENTIAL).

2144 AA; 233480 MW; 6EA898C1BA6555ECA CRC64;

Query Match 13.3%; Score 207.5; DB 1; Length 2144;

Best Local Similarity 23.6%; Pred. No. 4,7e-07;

Matches 82; Conservative 42; Mismatches 132; Indels 91; Gaps 16;

Qy 1 CERPCOHGATCMPAGEYE-----FQCLCRDGRKGLDCEHEENP 39

Db 457 CLREPCENYRCVSVLRFDSAFIASSVLFRIHPVGLRCRCPPGTFDTCETEVDL 516

Qy 40 COLRBPCLHGTCQ----GTRCLCLPFGSPRCQOG--SGH---GIAESDWH-----LEGS 86

Db 517 CYSR-PCGPHGHCRRREGGYTCLCRDGYTGEHCHEVSARGCRTPGVCKNGGTCVNLVGG 575

Qy 87 GGNDAP-GQYGAYFHDGDFLAPGHVF--SRSLPE-VPEITIEVLTSTASGILLWQGYE 142

Db 576 FKDCPSGDFEKFPCQVTRSPFARSFITRGLRQRFHFTLALSFAFKERDGLLLYN--- 632

Qy 143 VGEAGQKDFISLGLQDGHVFRYQIGSGEARLVSDP--INDGWHRY----- 189

Db 633 -GRNEKHDFVALEVIQEQVLTFSAGESTTTPVPGVSDGQWHHTQLKYNKPLIG 691

Qy 190 -TALREG---RRGSIQVDGEE-----LVSGRSPGNVANAKGSVYIG 828

Db 692 QTGLPQSPSEQVAVVSVVDCDVTGVALRFGAMLGNYSCAAQQTGGSKSLDTGTGPLL 751

```

QY 229 GAPPVATLTGGRFSSGTCVKNVLVHSAR-----PCAP 262
Db 752 GVFDLPSPFVVRMRH-FVCGMKNLQVDRHVDMAFDIANNCTVFCGP 797

RESULT 15
LMA2 MOUSE
ID LMA2 MOUSE STANDARD; PRT; 3106 AA.
AC Q60675; Q05003; Q64061;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Laminin alpha-2 chain precursor (Laminin M chain) (Merosin heavy chain).
GN LAMA2.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=FVB/N; TISSUE=Embryo, and Heart;
RX MEDLINE=95316259; PubMed=7795883;
RA Bernier S.M., Utani A., Sugiyama S., Doi T., Polistina C., Yamada Y.;
RT "Cloning and expression of laminin alpha 2 chain (M-chain) in the mouse."
RL Matrix Biol. 14:447-455(1995).
RN [2]
RP SEQUENCE OF 2162-2279 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Thymus;
RX MEDLINE=93346725; PubMed=8345183;
RA Chang A.C., Wadsworth S., Colligan J.E.;
RT "Expression of merosin in the thymus and its interaction with thymocytes."
RL J. Immunol. 151:1789-1801(1993).
RN [3]
RP SEQUENCE OF 64-281 FROM N.A.
RX MEDLINE=95179178; PubMed=7874173;
RA Xu H., Wu X.R., Wewer U.M., Engvall E.;
RT "Murine muscular dystrophy caused by a mutation in the laminin alpha 2 (Lama2) gene."
RL Nat. Genet. 8:297-302(1994).
RN [4]
RP SEQUENCE OF 20-25.
RX MEDLINE=21818471; PubMed=11829758;
RA Garbe J.H., Gohring W., Mann K., Timpl R., Sasaki T.;
RT "Complete sequence, recombinant analysis and binding to laminins and sulphated ligands of the N-terminal domains of laminin alpha3B and alpha5 chains."
RL Biochem. J. 362:213-221(2002).
RN [5]
RP BINDING TO FBLN1, FBLN2, AND NID2.
RX MEDLINE=99146904; PubMed=10022829;
RA Talts J.F., Andac Z., Goehring W., Brancaccio A., Timpl R.;
RT "Binding of the G domains of laminin alpha1 and alpha2 chains and perlecan to heparin, sulfatides, alpha-dystroglycan and several extracellular matrix proteins."
RL EMBO J. 18:863-870(1999).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 2932-3106.
RX MEDLINE=20085745; PubMed=10619025;
RA Hohenester E., Tisi D., Talts J.F., Timpl R.;
RT "The crystal structure of a laminin G-like module reveals the molecular basis of alpha-dystroglycan binding to laminins, perlecan, and agrin."
RL Cell 114:783-792(1999).
RN [7]
CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.
CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three

```

```

CC different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end. The alpha-2 chain is a subunit of laminin-2 (Merosin) and laminin-4 (S-merosin). Interacts with FBLN1, FBLN2 and NID2.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: Found in the basement membranes (major component).
CC -1- DOMAIN: The alpha-helical domains I and II are thought to interact with other laminin chains to form a coiled coil structure.
CC -1- DOMAIN: Domains VI, IV and G are globular.
CC -1- DISEASE: Defects in LAMA2 are a cause of murine muscular dystrophy (dy2J).
CC -1- SIMILARITY: Contains 1 laminin N-terminal domain.
CC -1- SIMILARITY: Contains 17 laminin EGF-like domains.
CC -1- SIMILARITY: Contains 5 laminin G-like domains.
CC -1- SIMILARITY: Contains 5 laminin G-like domains.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U12147; AAC52165.1; -
DR EMBL; X69869; CAA49502.1; -
DR EMBL; S75315; AAB33573.1; -
DR FAY; I49077; S53868.
DR PDB; 1QU0; 03-DEC-99.
DR PDB; 1DYK; 04-FEB-01.
DR MGD; MGI:99912; Lama2.
DR GO; GO:0005604; C:basement membrane; IDA.
DR InterPro; IPR008985; ConA like lec_gl.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR008979; Gal_bind like.
DR InterPro; IPR000034; Laminin B.
DR InterPro; IPR02049; Laminin EGF.
DR InterPro; IPR001791; Laminin G.
DR InterPro; IPR008211; LamNT.
DR Pfam; PF00052; Laminin B; 2.
DR Pfam; PF00053; Laminin_G; 14.
DR Pfam; PF00054; Laminin_G; 5.
DR PRINTS; PR00011; EGFLAMININ.
DR ProDom; PD003031; Laminin B; 1.
DR SMART; SM00180; EGF Lam; 14.
DR SMART; SM00281; LamB; 2.
DR SMART; SM00282; LamG; 5.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; 11.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01248; LAMININ TYPE EGF; 14.
DR PROSITE; PS00025; LAM G DOMAIN; 5.
DR KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 3106 LAMININ ALPHA-2 CHAIN.
FT DOMAIN 20 282 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 283 339 LAMININ EGF-LIKE 1.
FT DOMAIN 340 409 LAMININ EGF-LIKE 2.
FT DOMAIN 410 464 LAMININ EGF-LIKE 3.
FT DOMAIN 465 513 LAMININ EGF-LIKE 4.
FT DOMAIN 514 523 LAMININ EGF-LIKE 5 (N-TERMINAL)
FT DOMAIN 524 719 LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT DOMAIN 720 752 LAMININ EGF-LIKE 5 (C-TERMINAL).
FT DOMAIN 753 802 LAMININ EGF-LIKE 6.
FT DOMAIN 803 860 LAMININ EGF-LIKE 7.
FT DOMAIN 861 913 LAMININ EGF-LIKE 8.
FT DOMAIN 914 962 LAMININ EGF-LIKE 9.
FT DOMAIN 963 1009 LAMININ EGF-LIKE 10.
FT DOMAIN 1010 1055 LAMININ EGF-LIKE 11.

```

Db 2839 NDGQWHKIVRVKQEGILYVDD---ASSQTISPKKADILDVGGILYVGGLPINVTTRRI 2895
 Qy 239 GRFSSGITGCYKTNLVLHRSARFGAPPQPPLDL 269
 Db 2896 GPVYISLQGCVRNLRHEQA-----PVDDL 2918

Search completed: March 9, 2004, 17:19:44
 Job time : 7.86906 secs

FT DOMAIN 1056 1101 LAMININ EGF-LIKE 12.
 FT DOMAIN 1102 1161 LAMININ EGF-LIKE 13.
 FT DOMAIN 1162 1171 LAMININ EGF-LIKE 14 (N-TERMINAL).
 FT DOMAIN 1172 1375 LAMININ DOMAIN IV 2 (DOMAIN IV A).
 FT DOMAIN 1376 1415 LAMININ EGF-LIKE 14 (C-TERMINAL).
 FT DOMAIN 1416 1464 LAMININ EGF-LIKE 15.
 FT DOMAIN 1465 1522 LAMININ EGF-LIKE 16.
 FT DOMAIN 1523 1569 LAMININ EGF-LIKE 17.
 FT DOMAIN 1570 2140 LAMININ G-LIKE 1.
 FT DOMAIN 2141 2324 LAMININ G-LIKE 2.
 FT DOMAIN 2336 2517 LAMININ G-LIKE 3.
 FT DOMAIN 2522 2706 LAMININ G-LIKE 4.
 FT DOMAIN 2759 2930 LAMININ G-LIKE 5.
 FT DOMAIN 2929 3106 LAMININ G-LIKE 5.
 FT DOMAIN 1662 1863 COILED COIL (POTENTIAL).
 FT DOMAIN 1923 2146 COILED COIL (POTENTIAL).
 FT DISULFID 283 292 BY SIMILARITY.
 FT DISULFID 285 303 BY SIMILARITY.
 FT DISULFID 305 314 BY SIMILARITY.
 FT DISULFID 317 337 BY SIMILARITY.
 FT DISULFID 340 349 BY SIMILARITY.
 FT DISULFID 342 374 BY SIMILARITY.
 FT DISULFID 377 386 BY SIMILARITY.
 FT DISULFID 389 407 BY SIMILARITY.
 FT DISULFID 410 422 BY SIMILARITY.
 FT DISULFID 412 438 BY SIMILARITY.
 FT DISULFID 440 449 BY SIMILARITY.
 FT DISULFID 452 462 BY SIMILARITY.
 FT DISULFID 465 478 BY SIMILARITY.
 FT DISULFID 467 482 BY SIMILARITY.
 FT DISULFID 484 493 BY SIMILARITY.
 FT DISULFID 496 511 BY SIMILARITY.
 FT DISULFID 753 762 BY SIMILARITY.
 FT DISULFID 755 769 BY SIMILARITY.
 FT DISULFID 772 781 BY SIMILARITY.
 FT DISULFID 784 800 BY SIMILARITY.
 FT DISULFID 803 818 BY SIMILARITY.
 FT DISULFID 805 828 BY SIMILARITY.
 FT DISULFID 831 840 BY SIMILARITY.
 FT DISULFID 843 858 BY SIMILARITY.
 FT DISULFID 861 875 BY SIMILARITY.
 FT DISULFID 882 882 BY SIMILARITY.
 FT DISULFID 885 894 BY SIMILARITY.
 FT DISULFID 897 911 BY SIMILARITY.
 FT DISULFID 914 926 BY SIMILARITY.
 FT DISULFID 916 933 BY SIMILARITY.
 FT DISULFID 935 944 BY SIMILARITY.
 FT DISULFID 947 960 BY SIMILARITY.
 FT DISULFID 963 975 BY SIMILARITY.
 FT DISULFID 965 981 BY SIMILARITY.
 FT DISULFID 983 992 BY SIMILARITY.
 FT DISULFID 995 1007 BY SIMILARITY.
 FT DISULFID 1010 1019 BY SIMILARITY.
 FT DISULFID 1012 1026 BY SIMILARITY.
 FT DISULFID 1028 1037 BY SIMILARITY.
 FT DISULFID 1040 1053 BY SIMILARITY.
 FT DISULFID 1056 1068 BY SIMILARITY.
 FT DISULFID 1058 1075 BY SIMILARITY.
 FT DISULFID 1077 1086 BY SIMILARITY.
 FT DISULFID 1089 1099 BY SIMILARITY.
 FT DISULFID 1416 1425 BY SIMILARITY.
 FT DISULFID 1418 1432 BY SIMILARITY.

Query Match 13.3%; Score 207.5; DB 1; Length 3106;
 Best Local Similarity 35.8%; Pred. No. 7e-07;
 Matches 54; Conservative 25; Mismatches 53; Indels 19; Gaps 6;
 Qy 122 TIELEVRTSTASGILLMQGVEVAGQCKDFISLGLQDHLVFRYQIGSGEARLWSEDDI 181
 Db 2784 TIELEVRTAESGLLFYMG-RINHA----DFGTVQLRANGPFFSYDLGSGSTRMIPTKI 2838
 Qy 182 NDGEWHRVTAALREGRRGSIQVGDGEEVLVSGRSPGNVA--VNAKGSVYVIGAP-DVATLIG 238

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2004, 17:15:08 ; Search time 24.8707 Seconds
(without alignments)
3602.917 Million cell updates/sec

Title: US-10-006-011a-9
Perfect score: 1566
Sequence: 1 CERQPCGHGATCMPAGEYEF.....QPLDLQHRAGAGANTRPCPS 284

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPREMBL_25.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp Vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_virus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	357.5	22.8	2026	4	000468
2	349	22.3	925	5	Q9U7E8
3	346.5	22.1	488	4	Q96IC1
4	343	21.9	463	4	Q8NAL2
5	343	21.9	775	4	Q8N7Y0
6	325	20.8	295	11	Q80WX4
7	325	20.8	1009	11	Q8BGP3
8	304	19.4	283	11	Q8OV56
9	304	19.4	283	11	Q99KT4
10	304	19.4	350	11	Q8K326
11	295	18.8	478	5	Q86SD6
12	292	18.6	1361	5	Q9V7I4
13	292	18.6	1361	5	Q9NGV2
14	292	18.6	1534	4	Q75093
15	292	18.6	1618	4	Q9ULL7
16	283.5	18.1	1531	11	O88279

17	281	17.9	708	5	Q9NFS9
18	281	17.9	1035	5	Q9NEG1
19	281	17.9	4117	5	Q8IRV9
20	281	17.9	4179	5	Q9W4Y4
21	281	17.9	4223	5	Q8MPN3
22	281	17.9	4228	5	Q8IRV8
23	280.5	17.9	738	13	Q90Z45
24	280.5	17.9	1458	11	Q8COG8
25	280.5	17.9	1531	11	Q9WVB5
26	280.5	17.9	1557	11	Q80TR4
27	279.5	17.8	68	11	O08591
28	278.5	17.8	152	4	Q8N124
29	276	17.6	152	4	Q8N197
30	274.5	17.5	1530	11	Q9WUG5
31	270.5	17.3	1474	11	Q8COG5
32	265	16.9	181	4	Q8N4J5
33	264	16.9	4307	5	Q19319
34	258.5	16.5	1529	13	Q7ZXI2
35	250.5	16.0	1039	5	Q9VQB1
36	249	15.9	1410	5	Q20204
37	245.5	15.7	1515	13	Q9DE37
38	246	15.7	1523	11	O88280
39	244.5	15.6	1512	13	Q9DE36
40	244	15.6	1530	13	Q90WZ3
41	243.5	15.5	1025	11	Q9Z166
42	243.5	15.5	1521	11	Q9R1B9
43	241.5	15.4	783	13	Q90XG2
44	241.5	15.4	950	13	Q90Z44
45	241	15.4	1523	4	O75094

ALIGNMENTS

RESULT 1
O00468 AC O00468 PRELIMINARY; PRT; 2026 AA.
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE AGRIN precursor (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Buthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Groffen A.J.A., Buskens C.A.F., Van Kuppevelt T.H.M.S.M.,
RA Veerkamp J.H., Mommens L.A.H., van den Heuvel L.P.W.J.;
RL Eur. J. Biochem. 0:0-0(1998).
RN [2]
RP SEQUENCE OF 1-153 FROM N.A.
RX MEDLINE=96224170; PubMed=8617505;
RA Lennon G., Auffray C., Polymeropoulos M., Soares M.B.;
RT "The I.M.A.G.E. Consortium: an integrated molecular analysis of
RT genomes and their expression."
RL Genomics 33:151-152(1996).
RN [3]
RP SEQUENCE OF 1-153 FROM N.A.
RA Denzer A.J., Brandenberger R., Gesemann M., Chiquet M., Ruegg M.A.;
RL J. Cell Biol. 0:0-0(0).
DR EMBL; AF016903; AAC39776.1; -.
DR EMBL; U84406; AAB52917.1; -.
DR HSSP; P00740; IEDM.
DR Genew; HGNC:329; AGRN.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR004850; Agrin NtA.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003645; FcN.
DR InterPro; IPR002350; kazal.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.

DR InterPro; IPR00082; SEA domain.
 DR InterPro; IPR008993; UMP_like.
 DR InterPro; IPR001455; UMP0033.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00050; Kazal; 9.
 DR Pfam; PF00053; laminin_EGF; 2.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF03146; NEA; 1.
 DR Pfam; PF01390; SEA; 1.
 DR PRINTS; PRO0011; EGF_LAMININ.
 DR SMART; SM00180; EGF_Lam; 2.
 DR SMART; SM00274; FOLN; 5.
 DR SMART; SM00280; KAZAL; 9.
 DR SMART; SM00282; LamG; 3.
 DR SMART; SM00200; SEA; 1.
 DR PROSITE; PS00022; EGF_1; 6.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
 DR PROSITE; PS00024; SEA; 1.
 DR PROSITE; PS01148; UPP0033; 1.
 KW EGF-like domain; Laminin EGF-like domain; Signal.
 FT NON_TER 1
 FT SIGNAL <1 10 POTENTIAL.
 FT CHAIN 11 2026 AGRIN.
 FT SEQUENCE 2026 AA; 212881 MW; 4AB0EE710CD4B8EF CRC64;

Query Match 22.8%; Score 357.5; DB 4; Length 2026;
 Best Local Similarity 35.4%; Pred No. 4.7e-19;
 Matches 91; Conservative 29; Mismatches 110; Indels 27; Gaps 8;

QY 1 CERQPCQHGATMPAGEYFQCLRDGFKGDLCEHEENPCQLREPFCLHGFTQC-----GT 55
 1534 CLPNPCHGAPQCNLEAGRFHCQCPGRVGTCADEKSPCQ-PNFCGAAPCRVLPKGG 1592
 QY 56 RCLCLPGFSGPRCCQGGSGHGAESDWHLEGGGNDAPQCYAYFDDGFLAAPP-HVFSR 114
 1593 QCECLPREGFTCPQAS-----GQDGGPFADFNFSHLELRGLHTFAR 1637
 QY 115 SLPEVPTIELEVRTSTASGLLWQVVEGAGQKDFISLGLQDGHVFRYQLGSGEAR 174
 1638 DLGE-KMALEVFLARGPGLLYNGKTD--GKG-DFVSLARDRLRLEFRYDLGKGA 1693
 QY 175 LVSEDPINDGHWRTALREGRRSIQYDGEELVSGRSPGNVAVNAKGVYICGAPDVA 234
 1694 IRSREPVTLGAWTRVSLERNRKGALRVGDGPRVLGSPVPHVTLNKLKELYVGGAPDFS 1753
 QY 235 TLT-GGRFSSGTTGCVK 250
 1754 KLARAAYVSSGFDGAIQ 1770

RESULT 2
 Q9U7E8 ID Q9U7E8 PRELIMINARY; PRT; 925 AA.
 AC Q9U7E8
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE UNC-52/perlecan (Fragment).
 GN UNC-52.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=N2;
 RX MEDLINE=99443780; PubMed=10512861;
 RA Mullen G.P., Rogalski T.M., Bush J.A., Gortj P.R., Moerman D.G.;
 RT "Complex patterns of alternative splicing mediate the spatial and
 temporal distribution of Parlecan/UNC-52 in caenorhabditis elegans.";
 RL Mol. Biol. Cell 10:3205-3221(1999).

DR EMBL; AFL32983; AAD25092.1; -.
 DR InterPro; IPR008985; ConA_like_lec_gl.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00008; EGF; 3.
 DR Pfam; PF00047; Ig; 1.
 DR Pfam; PF00054; laminin_G; 2.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00408; IGC2; 1.
 DR SMART; SM00282; LamG; 3.
 DR PROSITE; PS00022; EGF_1; 3.
 DR PROSITE; PS00835; IGLIKE; 1.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
 KW EGF-like domain; Immunoglobulin domain.
 FT NON_TER 1
 FT SEQUENCE 925 AA; 102174 MW; 474BB5F045D67E0B CRC64;

Query Match 22.3%; Score 349; DB 5; Length 925;
 Best Local Similarity 31.2%; Pred. No. 8.1e-19;
 Matches 81; Conservative 41; Mismatches 102; Indels 36; Gaps 6;

QY 30 GDLCEHEE--NPCQLREPC LHGGTCQGTR-----CLCLPFGFSGPRCCQGGSGHGAESDWH 82
 679 GDVYSTQEPNNIC-ANSTCGMNGQCVRNMTHVTCCKLYYDGTCSLFFK----- 727
 Db 83 LEGSGNDAPQCYAYFDDGFLAAPPCHVPSRSLSEVPTIELEVRTSTASGLLWQVVE 142
 728 -----PIEHAARFDGDAFIELSSSEFFPHLTSEKDEIVAFKFKTEQQNGVLLWQGG-Q 777
 QY 143 VGEAGQKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGHWRTALREGRRSIQV 202
 778 RPTVQQMEYISVGI VNGHLHFSYELGGAAHLISEERVDDGKHSVRFERKGRGSGMRI 837
 QY 203 DEBELVSGRSPGNVAVNAKGVYICGAPDVAITLGGRFSSGTTGCVKVLHLSARPGAP 262
 838 DNYREYDGRSTGILAMLVNVDGNIFVYGGVVDISKATGGLFNNFVGCADVELNGVK---- 893
 QY 263 PPOPDLQHRAGANTRPC 282
 894 ----LDLMATAIDGRNVKFC 909

RESULT 3
 Q96IC1 ID Q96IC1 PRELIMINARY; PRT; 488 AA.
 AC Q96IC1
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RC Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC007649; AAH07649.1; -.
 DR InterPro; IPR008985; ConA_like_lec_gl.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR00210; IEGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00008; EGF; 3.
 DR Pfam; PF00054; laminin_G; 2.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00282; LamG; 2.
 DR PROSITE; PS00022; EGF_1; 3.
 DR PROSITE; PS01186; EGF_2; 1.

DR PROSITE; PS50025; LAM_G_DOMAIN; 2.
 KW Hypothetical protein; EGF-like domain.
 FT NON_TER 1
 SQ SEQUENCE 488 AA; 51764 MW; A732B99CC680E772 CRC64;

Query Match 22.1%; Score 346.5; DB 4; Length 488;
 Best Local Similarity 35.3%; Pred. No. 5.6e-19;
 Matches 89; Conservative 29; Mismatches 107; Indels 27; Gaps 8;

QY 6 CQHATCMFAGYERQCLCEDGFKDLCHEENPCOLREPCLHGTCQ-----GTRCLCL 60
 DB 1 CHGGAPQCNLEGRFHCQPPRGVPTCADEKSPQ--PNPCHGAAPCRVLEGGQCECP 59
 QY 61 PFGSPRCQQGGHGHGTAESDWHLEGGNDAPFGQYGFHDDGFLAFFG-HVFSRSLPEV 119
 DB 60 LGREGFTFCQTAS-----GQDGGPFLADFNQSPHLELRGLRHTFARDLGE- 103
 QY 120 PETIELEVTSTASGLLLWQGVVEGAGQCKDFISLGLQDGHVFRYQLGSGEARLVSED 179
 DB 104 KMALEVYVFLARGPSSGLLYNGQKTD--GKG-DFVSLALRDRLEFRYDLGKGAAVIRSE 160
 QY 180 PINDGEHVRVTLRGRRGSIQVDSBELVSGRSPGNVAVNAKSGVYIGGAPDVATLT-G 238
 DB 161 PVLGATWTRVSLERNGKALRVGDGPRVLGSSPVPHITVNLKEPLYVGGAPDFSKLARA 220
 QY 239 GRPSSGITGVCK 250
 DB 221 AAVSSGFDGAIQ 232

RESULT 4
 Q8NAL2 PRELIMINARY; PRT; 463 AA.
 AC Q8NAL2;
 DT 01-OCT-2002 (T-EMBLrel. 22, Created)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Hypothetical protein FLJ35160.
 OS Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
 RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
 RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isegai T.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK092479; BAC03900.1;
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR InterPro; IPR001881; EGF_CA.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00054; laminin_G; 2.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SMO0282; LamC; 2.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 2.
 KW Hypothetical protein; EGF-like domain.
 SQ SEQUENCE 463 AA; 50635 MW; F688BF2714D5D08C CRC64;

Query Match 21.9%; Score 343; DB 4; Length 463;
 Best Local Similarity 33.5%; Pred. No. 9.9e-19;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

QY 34 EHEENPCOLREPCLHGTC-----QGRCLCLPFGSPRCQQGGHGHGTAESDWHLEGGGN 89
 DB 236 ENAAHPC-VRAPCAAGSGSCRPEKGYDCDCPLGFEGLHCOKAIEAI----- 281
 QY 90 DAP-----GQYGFHDDGFLAFFGHVFSRSLPEVPEVETIELEVTSTASGLLLWQGVVEGE 145
 DB 282 EIFQIGRSYLTVDNPDILKRVSG-----SRK-----NVFMRFKTKTKADGGLLLWRG--DSP 330
 QY 146 AQGQKDFISLGLQDGHVFRYQLGSGEARLVSEDPINDGEHVRVTLRGRRGSIQVDSG 205
 DB 331 MRPNSDFISLGLRDGALVFSYNLGSGVASIMVNGSFPDGRHVRKAVRDGQSGKITVDDY 390
 QY 206 ELVSGRSPGNVAVNAKSGVYIGGAPDVATLTGCRSSGITGVCKNLVLSARPPQP 265
 DB 391 GAETGKSPGMRQLNICALYVGGMKIEALHTRQYMRGLVGCISHFTLST-----DY 443
 QY 266 PLDIQRAQAQANTRPC 282
 DB 444 HISLVEDAVDQKINTC 460

RESULT 5
 Q8N7Y0 PRELIMINARY; PRT; 775 AA.
 AC Q8N7Y0;
 DT 01-OCT-2002 (T-EMBLrel. 22, Created)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Hypothetical protein FLJ40230.
 OS Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
 RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,
 RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isegai T.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK097549; BAC05096.1;
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR InterPro; IPR000985; Ccna_like lec_3.
 DR InterPro; IPR001881; EGF_CA.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00054; laminin_G; 3.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SMO0282; LamC; 3.
 DR PROSITE; PS00022; EGF_1; 3.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 3.
 KW Hypothetical protein; EGF-like domain.
 SQ SEQUENCE 775 AA; 84800 MW; 61A81294F0204ED2 CRC64;

Query Match 21.9%; Score 343; DB 4; Length 775;
 Best Local Similarity 33.5%; Pred. No. 1.9e-18;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;

QY 34 EHEENPCOLREPCLHGTC-----QGRCLCLPFGSPRCQQGGHGHGTAESDWHLEGGGN 89
 DB 548 ENAAHPC-VRAPCAAGSGSCRPEKGYDCDCPLGFEGLHCOKAIEAI----- 593
 QY 90 DAP-----GQYGFHDDGFLAFFGHVFSRSLPEVPEVETIELEVTSTASGLLLWQGVVEGE 145

Query Match 21.9%; Score 343; DB 4; Length 463;
 Best Local Similarity 33.5%; Pred. No. 9.9e-19;
 Matches 86; Conservative 34; Mismatches 97; Indels 40; Gaps 8;